Chasing the Tail: The Emergence of Autocatalytic Networks

Wim Hordijk[a], Mike Steel[b]

[a]Konrad Lorenz Institute for Evolution and Cognition Research, Klosterneuburg, Austria
[b]Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand

Abstract

A ubiquitous feature of all living systems is their ability to sustain a biochemistry in which all reactions are coordinated by catalysts, and all reactants (along with the catalysts) are either produced by the system itself or are available from the environment. This led to the hypothesis that 'autocatalytic networks' play a key role in both the origin and the organization of life, which was first proposed in the early 1970s, and has been enriched in recent years by a combination of experimental studies and the application of mathematical and computational techniques. The latter have allowed a formalization and detailed analysis of such networks, by means of RAF theory. In this review, we describe the development of these ideas, from pioneering early work of Stuart Kauffman through to more recent theoretical and experimental studies. We conclude with some suggestions for future work.

Keywords: Autocatalytic sets, origin of life, binary polymer model, RNA world

1. The origin of life

In the early 17th century, the Flemish chemist Jan Baptist van Helmont wrote the following [50]:

> "If you press a piece of underwear soiled with sweat together with some wheat in an open mouth jar, after about 21 days the odor changes and the ferment coming out of the underwear and penetrating through the husks of the wheat, changes the wheat into mice."

This reflected the commonly held belief at that time, even among scientists, of *spontaneous generation*. Life arises spontaneously and continuously: mice from wheat, maggots from meat, frogs from mud, etc. Hence, the origin of life was not considered a scientific question.

It was not until more than two centuries later, in 1862, that Louis Pasteur won a price from the French Academy of Sciences for definitively putting to rest the idea of spontaneous generation. He performed a simple but clever experiment, showing that nothing happened to a sterilized broth contained in a flask in which dust particles could not reach the broth, but that micro-organisms quickly appeared in the broth after the curved neck of the flask was broken [50]. Pasteur thus concluded that all life comes from other life.

At around the same time, in 1859, Darwin published his now famous book *On the Origin of Species* [7]. One of the main ideas underlying his theory of evolution by natural selection is that of *common descent*: any (arbitrary) group of currently living species will, if you go far enough back in time, have a common ancestor. As a consequence, *all* life on earth must have come from one (or just a few) common ancestor(s).

So, if all life comes from life, going all the way back to a "last universal common ancestor" (LUCA), then where did this common ancestor, one of the very first living organisms, come from? The origin of life had become a genuine scientific problem.

Currently, the main paradigm in the origin of life field is that of an RNA world [15]. Given the common (and apparently ancient) functionality of RNA in the molecular machinery underlying all life as we know it [38], one of the earliest stages in (or towards) life is assumed to have existed exclusively of RNA molecules that were responsible for both the replication and expression of genetic information (through their catalytic properties). However, despite progress towards the experimental spontaneous formation of RNA [52], the RNA world hypothesis is not without problems [61, 4], and so far nobody has been able to show that RNA can catalyze its own template-directed replication.

What has been shown experimentally, though, is that some RNA molecules can efficiently catalyze the formation of *other* RNA molecules from shorter RNA fragments [35]. Moreover, there are experimentally constructed sets of RNA molecules that *mutually* catalyze each other's formation [54, 43, 62]. In other words, rather than having each RNA molecule replicate itself (a tall order), they all help each other's formation from their basic building blocks, in a network of molecular collaboration [18, 49].

2. Autocatalytic sets (RAFs)

Such a collaborative molecular network is an instance of an *autocatalytic set*, a concept that was originally introduced by Kauffman [39, 40, 41]. Initial (computational) investigations into such sets were done in the late 80s and early 90s [12, 2, 3], and the first experimental autocatalytic set was constructed in the lab in 1994, consisting of two complementary nucleotide-based oligomers [54]. Later on, the concept was made mathematically more rigorous and studied in more detail, both theoretically and computationally, through the development of RAF (Reflexively Autocatalytic and F-generated) theory [57, 22, 48, 20].

To understand the basic idea behind the concept of an autocatalytic set, consider the cross-catalytic set of oligomers of Sievers and von Kiedrowski [54], which is depicted in a chemical network representation in Figure 1 (reproduced from Patzke and von Kiedrowski [51]). The basic building blocks are the trimers **A** and **B**, which are each other's base-pair complement when read in opposite directions, indicated by the thick white arrows in Figure 1. The fully formed hexamers **AA** and **BB** now serve as templates to which the complementary trimers can bond through base-pairing (again, in the opposite direction, as in RNA double strand formation). For example, two **A** trimers can attach to a **BB** template, allowing them to ligate into a fully formed **AA** hexamer. After strand separation, the original **BB** template is regained, plus a new **AA** template. In a similar way, such an **AA** template can catalyze the formation of another **BB** template from two **B** trimers. In other words, the two oligomers **AA** and **BB** cross-catalyze each other's formation from their basic building blocks.
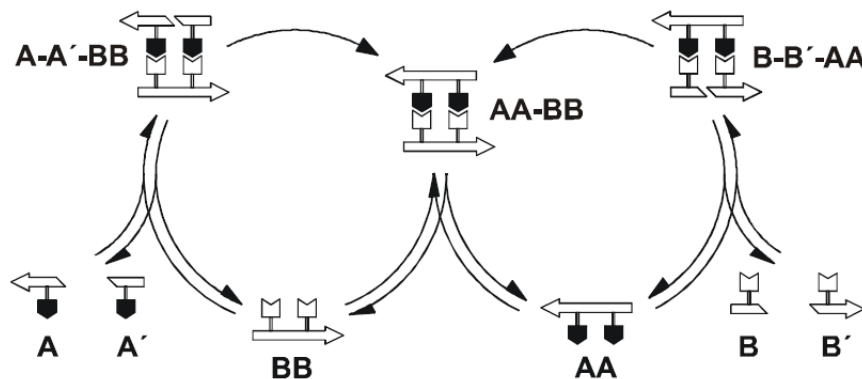

Figure 1: A chemical network representation of the nucleotide-based oligomer autocatalytic set of Sievers and von Kiedrowski [54]. Reproduced from Patzke and von Kiedrowski [51].

Sievers and von Kiedrowski [54] were able to create such a system experimentally using **A**=CCG and **B**=CGG. In practice, there were some hurdles to overcome, such as efficient strand separation, but it was the first experimental proof of principle of an autocatalytic set. Later on, a similar autocatalytic set of two much longer cross-catalytic RNA ligases (> 70 bp) was constructed experimentally by Kim and Joyce [43]. Moreover, these RNA ligases were subjected to

an artificial form of evolution, significantly increasing their catalytic efficiency [45].

Such a cross-catalytic RNA network can be represented more abstractly as shown in Figure 2, where round dots represent molecule types and square boxes represent reactions. Solid black arrows indicate reactants going into and products coming out of a reaction, and dashed gray arrows indicate catalysis. The network in Figure 2 represents an autocatalytic set consisting of two reactions where the two products mutually catalyze each other's formation from their basic building blocks (the four reactants). Note that none of the molecules in this network is a self-replicator, but the set as a whole is able to efficiently reproduce itself given a steady supply of building blocks.
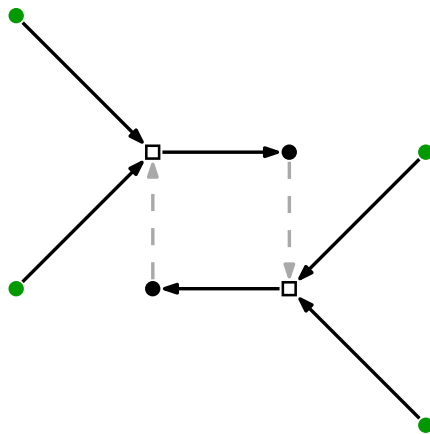

Figure 2: A more abstract representation of the oligomer autocatalytic set of Figure 1. Round dots represent molecule types (the four outer food molecules are indicated in green) and square boxes represent chemical reactions. Solid black arrows indicate reactants going into and products coming out of a reaction, and dashed grey arrows indicate catalysis.

Of course, this basic idea of an autocatalytic set can be generalized to any number of molecule types and reactions, and can be defined more formally. First, we define a *chemical reaction system* (CRS) as a tuple $\mathcal{Q} = (X, \mathcal{R}, C)$, where:

- X is a set of molecule types: $X = \{a, b, c, ...\}$,

- $\mathcal{R}$ is a set of reactions of the form $r_i = a + b + \ldots \rightarrow c + \ldots$,

- C is a set of catalytic assignments, indicating which molecule types catalyze which reactions: $C = \{(x_{i1}, r_{j1}), (x_{i2}, r_{j2}), \ldots\}$, where $x_i \in X$ and $r_j \in \mathcal{R}$ (thus C is a subset of the product set $X \times \mathcal{R}$).

In addition, we define a food set $F \subset X$ consisting of the basic building blocks of the CRS (i.e., those molecule types that can be assumed to be available from the environment). An autocatalytic (or RAF) set is now defined as a subset $\mathcal{R}' \subset \mathcal{R}$ of reactions (and the associated molecule types) which is:

1. *Reflexively Autocatalytic* (RA): each reaction $r \in \mathcal{R}'$ is catalyzed by at least one molecule type involved in $\mathcal{R}'$, and
2. *Food-generated* (F): all reactants used in reactions from $\mathcal{R}'$ can be created from the food set F by using a series of reactions only from $\mathcal{R}'$ itself.

Thus an autocatalytic set forms a catalytically closed (RA) and self-sustaining (F) reaction network. A more formal (mathematical) definition of RAF sets is provided in Hordijk and Steel [22] and Hordijk et al. [29]. The phrase "involved in $\mathcal{R}'$" in condition (RA) refers to a molecule type that is either a reactant or product of a reaction in $\mathcal{R}'$, or an element of the food set F.

The notion of a RAF set captures the two key components that are coupled in a living system: it is self-sustaining from resources available in the environment (the F-generated condition), and all reactions are coordinated by catalysts generated within the system itself (the RA condition). Note that catalysts not only accelerate biochemical reactions by several orders of magnitude, but they also allow for the synchronization of these reactions by greatly reducing the variance in reaction rates [66].

We will use the terms "autocatalytic set", "autocatalytic network", and "RAF set" interchangeably, with the latter usually referring to the mathematical formalization of the general concept.

In the oligomer example of Figure 1, we thus have:

- $X =\{\mathbf{A},\mathbf{B},\mathbf{AA},\mathbf{BB}\}$,

- $F =\{\mathbf{A},\mathbf{B}\}$,

- $\mathcal{R} =\{r_1=\mathbf{A}+\mathbf{A}\rightarrow\mathbf{AA},r_2=\mathbf{B}+\mathbf{B}\rightarrow\mathbf{BB}\}$,

- $C =\{(\mathbf{AA},r_2),(\mathbf{BB},r_1)\}$,

where $\mathcal{R}$ forms a RAF set. However, as already mentioned, a RAF set can be of any size. For example, Vaidya et al. [62] have constructed experimental autocatalytic networks of ribozymes (RNA catalysts) with up to 16 reactions (see Figure 3). Moreover, the definition of an autocatalytic set is not restricted to just RNA molecules. In principle, the molecule set X can consist of any kind of molecules, or even mixtures of molecule types. For example, Ashkenasy et al. [1] have created experimental autocatalytic sets of peptides, and Sousa et al. [56] have shown that the metabolic network of *Escherichia coli* contains a RAF set consisting of close to 1200 molecule types (of different kinds) and almost 1800 reactions. Clearly, autocatalytic sets are not just a theoretical concept, but they exist in real chemical and biological networks.

2.1. Related notions

The concept of a RAF needs to be distinguished from a weaker concept where condition (F) is replaced by "all reactants used in reactions in $\mathcal{R}'$ are products of reactions from $\mathcal{R}'$ or present in the food set F". This weaker notion is called a *pseudo-RAF*. As a model for the origin of biochemistry it is problematic since it cannot always form by starting with just the food set.
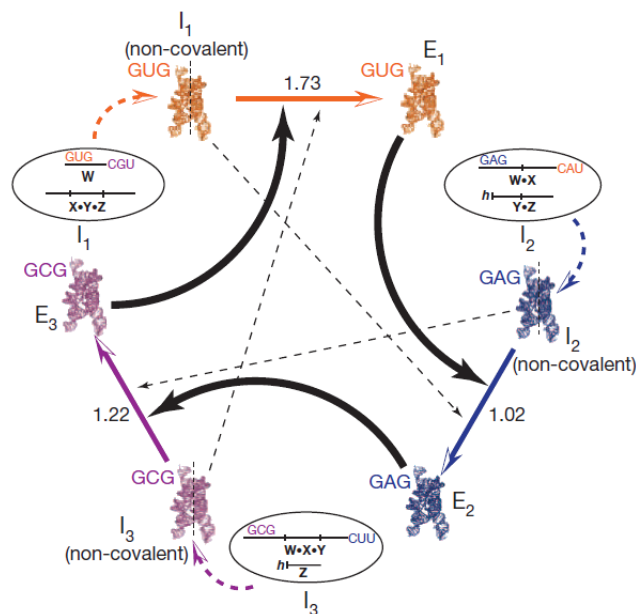
Figure 3: An example of a three-member autocatalytic RNA network. Similar autocatalytic networks with up to 16 members have been created in the lab. Reproduced from Vaidya et al. [62].

Notice also that the definition of a RAF means that some reactions may need to proceed uncatalyzed (at a slower rate) until a catalyst becomes available; for example, this is the case for the simple system shown in Fig. 2. Insisting that a reaction can only proceed if the catalyst is already available is a much more restrictive condition, and a RAF with this property is called a *constructively autocatalytic F-generated* (CAF) set. Such systems typically require unrealistically high rates of catalysis to form [48].

A further notion is that of a co-RAF, introduced in Steel et al. [59]. Given a CRS and food set, a *co-RAF* is a set of one or more reactions that can be added to an existing RAF to create a larger RAF (an example is shown later in Fig. 9). Notice that every RAF can be represented as the union of at least one minimal RAF (i.e., a RAF which does not contain another RAF as a strict subset) together with a co-RAF. This way of describing a RAF corresponds to the notion of a "viable core" and a "periphery" in Vasas et al. [64].

RAFs are related to, but different from other approaches to the formal description of metabolic networks, including (M, R) systems [53] and Chemical Organization Theory [8]. For details of the connections to these concepts (along with further details on RAFs, pseudo-RAFs and CAFs), see Steel et al. [59] and the references therein.

Finally, note that RAFs are more general than hypercycles. In a hypercycle [9, 46] each member helps in the *replication* of the next one in the cycle, whereas

in an autocatalytic set the members mutually help each other's *formation*. However, an autocatalytic set can also have members that are self-replicators, and a hypercycle is therefore a specific instance of an autocatalytic set, but they are not the same, and they are unfortunately often confused in the literature [60]. Moreover, there are no chemical examples of hypercycles (they only exist as mathematical models), whereas there are several experimental and biological examples of autocatalytic sets, as mentioned above.

3. Required levels of catalysis

A central question now arises: how likely is it that autocatalytic sets form spontaneously? To investigate this question, Kauffman [40, 41] introduced a simple and abstract model of a CRS, now known as the *binary polymer model*. In this model, molecules are represented by binary strings (i.e., sequences of zeros and ones) that can be concatenated into longer ones or cut into smaller ones. Catalysis is assigned at random. More specifically:

- The molecule set X consists of all binary strings up to (and including) a maximum length n. In mathematical notation: $X = \{0,1\}^{\leq n}$.

- The food set F consists of all binary strings up to (and including) length t, where $t \ll n$. In most cases, $t = 2$ is used (i.e., $F = \{0, 1, 00, 01, 10, 11\}$).

- The reaction set $\mathcal{R}$ consists of two types of reactions:

  1. *Ligation*: concatenating two binary strings together into a longer one, e.g., $00 + 111 \to 00111$.
  2. *Cleavage*: cutting a binary string into two smaller ones (e.g. $010101 \to 01 + 0101$).

  All possible ligation and cleavage reactions between binary strings are included in $\mathcal{R}$, as long as none of the reactants or products violates the maximum molecule length n. Note that each ligation reaction has a corresponding (reverse) cleavage reaction.

- The catalysis set C is constructed at random for each instance of the model, as follows. For each molecule type $x \in X$ and reaction $r \in \mathcal{R}$, the pair (x, r) is independently included in C with probability p.

Thus the model has two parameters (assuming we keep $t = 2$ fixed): the maximum molecule length n and the probability of catalysis p. Each instance of the model gives rise to a different CRS. However, note that for a given n, the molecule set X and the reaction set $\mathcal{R}$ are the same for each instance, but the catalysis set C will be different (because of the random catalysis assignments). Figure 4 shows an example of a RAF set that arose in an instance of the binary polymer model with $n = 5$ and $p = 0.0045$. We will discuss algorithms for detecting RAF sets in Section 5.
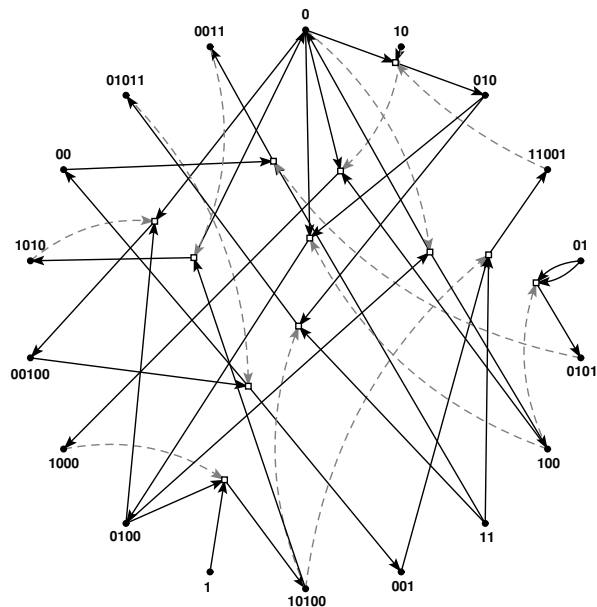
7

Figure 4: An example of a RAF set that arose in an instance of the binary polymer model with $n = 5$ and $p = 0.0045$. The molecule types (binary strings up to length five) are shown around a circle, with the reaction boxes placed randomly inside the circle. The food set consists of monomers and dimers.

Kauffman then claimed that in this particular model, autocatalytic sets are an expected emergent property. His argument was motivated, in part, by the analogous transitions that occur in the much-studied random graphs of the Erdős–Rényi type [10, 11]. Given a set of nodes, imagine that you randomly start to add directed edges and keep track of the existence of cycles, or add undirected edges and keep track of the largest connected component in the graph. Initially, there will be no cycles (in the directed case) and for the undirected graph, the largest component will be very small. However, if you keep adding (directed or undirected) edges to the graph, at some point, there will be a sudden transition to the emergence of cycles or, for undirected graphs, a large connected component.

Kauffman noted that these models are not directly applicable to the binary polymer model. However, he showed that similar arguments can be applied. Using some basic combinatorics, it can be shown that with increasing n in the binary polymer model, the number of molecule types (binary strings) increases as $\mathcal{O}(2^n)$ and the number of reactions increases as $\mathcal{O}(n \cdot 2^n)$. In other words, the number of reactions grows n times faster than the number of molecule types. In terms of our imaginary graph above, where the nodes now represent molecule

types and the edges represent catalyzed reactions, this means that the ratio of edges to nodes grows as $\mathcal{O}(n)$ with increasing n (if we keep p fixed, the number of *catalyzed* reactions grows at the same rate as the *total* number of reactions). Consequently, if you make n large enough, while keeping p fixed, autocatalytic networks will form (as a giant connected component). In other words, for a given probability of catalysis p, there is always a large enough maximum molecule length n for which autocatalytic sets are highly likely to exist in random instances of the binary polymer model.

Mathematically, this is a sound argument, but practically it could be problematic. As Lifson [44] correctly pointed out, for a fixed value of p with increasing n, the average number of reactions catalyzed per molecule increases exponentially. In the binary polymer model, the average number f of reactions catalyzed per molecule is simply $f = p|\mathcal{R}|$, and, as was stated above, $|\mathcal{R}|$ grows as $\mathcal{O}(n \cdot 2^n)$ with increasing n. So, depending on how large n needs to be to reach the giant connected component threshold, one could easily end up with each molecule having to catalyze an exponentially large number of reactions for autocatalytic sets to form. Chemically, this is highly implausible. Instead, as Lifson [44] argued, what should be kept constant is the average *level* of catalysis $f = p|\mathcal{R}|$, not the *probability* of catalysis p. In that case, it is not immediately clear whether autocatalytic sets are likely to form, even for very large n.

In fact, one of us (MS) proved mathematically that for a constant level of catalysis f, the probability of RAF sets existing in instances of the binary polymer model quickly converges to zero with increasing n. However, Steel [57] also proved mathematically that, unlike Kauffman's required exponential growth rate, a *quadratic* growth rate $\mathcal{O}(n^2)$ in the level of catalysis f is already sufficient for RAF sets to form with high probability for increasing n.

To further investigate a conjecture made by Steel [57], we then performed computer simulations to gain more insight into the probability of RAF sets forming in instances of the binary polymer model. After running on a large computer cluster for several weeks, these simulations suggested that even a *linear* growth rate in the level of catalysis f (with increasing n) may already be sufficient (see Figure 5). This suggestion, and thus the original conjecture, were subsequently verified theoretically [48], with the simulations indicating that only between one and two reactions being catalyzed per molecule (on average) suffices for autocatalytic sets to arise with high probability in the binary polymer model (at least for n up to 50, by extrapolation). Moreover, the restriction of the polymer model to binary sequences can be easily relaxed; all of the formulae generalize seamlessly to sequences over a k-letter alphabet for any $k \geq 2$.

Later on, we also investigated several variants of the standard binary polymer model, such as (i) using a template-matching rule to assign catalysts, rather than assigning them purely randomly [29]; (ii) considering a "partitioned" reaction network of two different classes of molecule types, where reactions can only happen between molecules from the same class, but catalysis can happen both within and between classes [55]; (iii) using only the longest binary strings as catalysts [33] or (iv) using a power law distribution for catalysis assignments [31]. In all cases, it turned out that the main result that a linear growth rate in
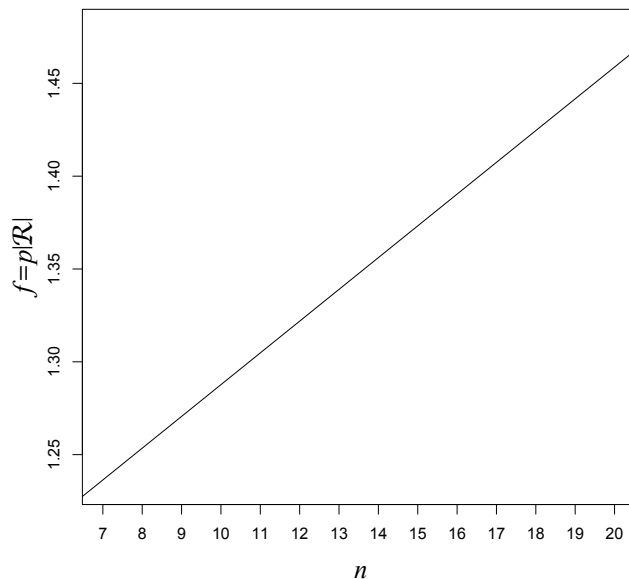
Figure 5: The average level of catalysis f required to get a 50% probability of RAF sets to form in random instances of the binary polymer model, for increasing values of n. The solid line is a least-squares regression fit to the data, and clearly suggests a linear relation.

the level of catalysis is sufficient, holds up. In some of these cases, it is actually possible to predict mathematically what the required level of catalysis is in the model variant based on the corresponding level from the standard model [23].

Moreover, we have now also shown theoretically that *any* distribution-based variant of the binary polymer model requires at most a linear growth rate in the level of catalysis (with increasing n) to generate autocatalytic sets form with high probability [28]. In other words, regardless of which probability distribution is used to determine the catalysis rate of each molecule type, the level of catalysis required for autocatalytic sets to form will never need to grow faster than linearly. So, this upper bound on the growth rate in the required level of catalysis is indeed a very robust one.

Finally, as Figure 5 shows for n up to 20 in the binary polymer model, each molecule needs to catalyze at most $f = 1.5$ reactions on average to get a high probability of RAF sets. Extrapolating the linear trend, even for $n = 50$, no more than $f = 2$ reactions being catalyzed per molecule, on average, is sufficient for autocatalytic sets to form. These numbers are chemically highly plausible, as it is known that many molecules, including proteins, can catalyze more than one reaction [36, 37]. In fact, in the 16-member RNA autocatalytic set of Vaidya et al. [62], each ribozyme catalyzes exactly four reactions, and

in the nine-member peptide autocatalytic network of Ashkenasy et al. [1], the average number of reactions catalyzed per peptide is about three.

4. The role of the food set

Recall that the definition of a RAF set contains two parts. As well as all reactions needing to be catalyzed by some molecule from the set itself, these molecules must also be produced from an appropriate food set using only reactions from the set itself. This "food-generated" constraint is an important one.

In a related and similar study, using a dynamical version of the binary polymer model, catalytic cycles were identified (as strongly connected components, or SCCs), but observed to be unstable [13]. In other words, they were able to occasionally form spontaneously, but not to sustain themselves. Indeed, these catalytic cycles (SCCs) only satisfied the first (RA) part of the RAF definition, but not the second (F) part. Hence, since they were not food-generated, they were not self-sustaining structures.

In a follow-up study, this distinction was made explicit and investigated in more detail [14]. Catalytic cycles (SCCs) already start showing up at much lower levels of catalysis (see Figure 6) but are unstable. Only by increasing the level of catalysis is it possible to get truly self-sustaining RAF sets to form. However, as was already shown above, this required level of catalysis is still well within a chemically plausible range.
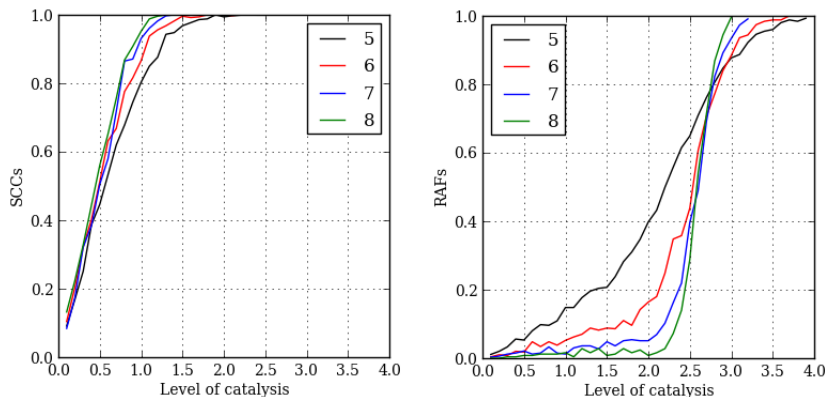

Figure 6: The probability of SCCs (left) and RAFs (right) for different maximum molecule lengths ($n = 5, 6, 7, 8$) and levels of catalysis ($f \in [0.0; 4.0]$) in instances of the binary polymer model using only forward (ligation) reactions. Reproduced from Filisetti et al. [14].

In fact, it can be shown mathematically that under the assumption that no food molecules catalyze any reactions, a RAF set always necessarily contains at least one catalytic cycle, but such a cycle is not necessarily a RAF set [59]. The assumption of no food molecules catalyzing any reaction is no serious constraint

though, as in the binary polymer model with $t = 2$, the food set becomes exponentially small for increasing n. In fact, the main results, as described in the previous section, do not change in any way when this restriction is explicitly taken into account.

In general, however, the size of the food set can play an important role. In the extreme case of $F = X$, every CRS is trivially food-generated. This brings up the question of the *minimum* food set for which a given RAF set still maintains its RAF properties. Unfortunately, finding such a minimum food set is an NP-complete problem and, moreover, it may not be unique [56]. This leads us to the algorithmic aspects of RAF sets.

5. Algorithmic aspects

In Hordijk and Steel [22], we introduced an algorithm for detecting RAF sets in any given reaction network which runs in polynomial time in the size of the CRS. A straightforward analysis showed that the worst-case running time of this RAF algorithm is $\mathcal{O}(|X||\mathcal{R}|^3)$. In Hordijk et al. [29, 34] several simplifications and improvements were introduced, reducing the worst-case running time to $\mathcal{O}(|\mathcal{R}|^2 \log |\mathcal{R}|)$, and giving even more of a speed-up in certain specific cases. Moreover, it was shown that the actual (average) running time of the RAF algorithm on random instances of the binary polymer model is actually sub-quadratic [22].

So, finding RAF sets is easy. In fact, if a given CRS contains a RAF, the RAF algorithm returns the *maximal* RAF set (maxRAF) and consists of the union of all RAF (sub)sets that exist within that CRS. If a CRS does not contain a RAF, the RAF algorithm simply returns an empty set. Given the existence of a RAF set, the algorithm can then be used recursively to find RAF subsets.

A minimal RAF set (i.e., an RAF set that does not contain a proper subset that is also a RAF) is called an *irreducible* RAF set (irrRAF). Notice that any smallest RAF in a CRS is necessarily an irrRAF, but larger irrRAFs may also exist in a CRS. In Hordijk et al. [30] we showed that, in principle, there can be exponentially many irrRAFs within a maxRAF. Thus, enumerating *all* irrRAFs is an NP-complete problem. Moreover, finding a *smallest* irrRAF (which is not necessarily unique) is NP-complete, and so is even just determining the *size* of such a smallest irrRAF. However, irrRAFs can be sampled using a randomized extension of the RAF algorithm (see Figure 7), so an approximation of the irrRAF size distribution and the smallest irrRAF size can be obtained in polynomial time [59].

In biological systems, inhibitors (as well as catalysts) also play an important role in regulation. The notion of inhibition can be included in the RAF framework as well. Next to requiring that each reaction in the RAF set is catalyzed by at least one of the molecules from the set itself, it is now also required that none of the reactions are *inhibited* by any of the molecules from the same set. Unfortunately, finding such "uninhibited" RAFs (or *u*-RAFs) is also NP-complete [48]. In Hordijk et al. [34], we introduced a fixed-parameter tractable algorithm for finding *u*-RAFs when the total number m of inhibitors
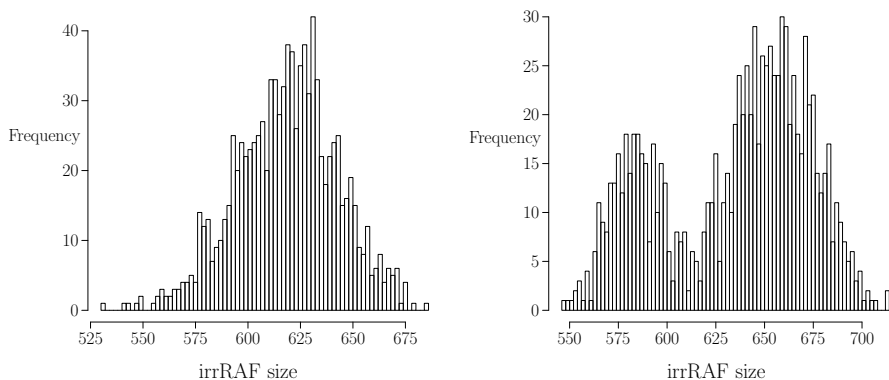
Figure 7: Histograms of irrRAF sizes from samples of size 1000 from two different maxRAFs found in instances of the binary polymer model with $n = 10$ and $p = 0.000075$. Reproduced from Steel et al. [59].

in the system is limited. If m is small enough, the problem can still be solved efficiently, as we showed with an application to the binary polymer model with inhibition included, and using $m = 10$ [34].

In Hordijk and Steel [28], we then introduced an efficient but approximate algorithm for finding u-RAFs, independent of the value of m. This algorithm is based on a "sufficient condition" for a u-RAF to exist. In other words, if the algorithm returns a non-empty result, the given CRS contains a u-RAF, although the maximal u-RAF may be larger than the u-RAF found by the algorithm. However, if the algorithm returns an empty set, the given CRS may still contain one or more u-RAFs. The performance of this algorithm was then also tested on instances of the binary polymer model with inhibition included [28]. In this way, u-RAFs were detected at certain levels of inhibition that were *higher* than those for catalysis (as predicted by a mathematical analysis that suggested that such u-RAFs should exist within a narrow range of rates).

Several other RAF-related problems also turn out to be NP-complete, such as finding a minimum food set for which a given RAF still exists, as already mentioned above. The challenge remains to construct efficient and effective approximation algorithms for some of these difficult problems.

6. The structure and evolvability of RAF sets

As was explained in Section 3, Kauffman's argument to support his claim of a high likelihood of autocatalytic sets existing in instances of the binary polymer model was analogous to the phase transition in random graphs where giant connected components appear. Because of this argument, autocatalytic sets have been criticized for lacking evolvability [63]. In this viewpoint, an autocatalytic set either exists, as a gigantic connected component, or it does not. And even if it exists, there is little room for change and adaptation, as it already consists of almost the entire reaction network.

However, at catalysis levels where RAFs first appear in the binary polymer model, they generally involve only a small proportion (around 10%) of all possible reactions (though they are not small either [59]). Moreover, as we have shown, RAF sets often consist of an entire hierarchy of smaller and smaller RAF subsets (subRAFs), with the irrRAFs residing at the bottom of this hierarchy [30]. Although a CRS that has a RAF always has a unique maximal RAF, it may have many subRAFs which form a partially ordered set (poset) under set inclusion. For any CRS (not just the binary polymer model), there is an algorithm that constructs the poset of subRAFs in polynomial time in the joint sizes of the CRS and the poset (note, however, that the poset itself can be exponentially large in the size of the CRS).

Figure 8 illustrates a maxRAF consisting of five reactions $r_1, \ldots, r_5$ from an instance of the binary polymer model, along with its poset of subRAFs, represented by its Hasse diagram, in which there is an arc from each RAF down to any maximal subRAF it contains. Note that there are two irrRAFs
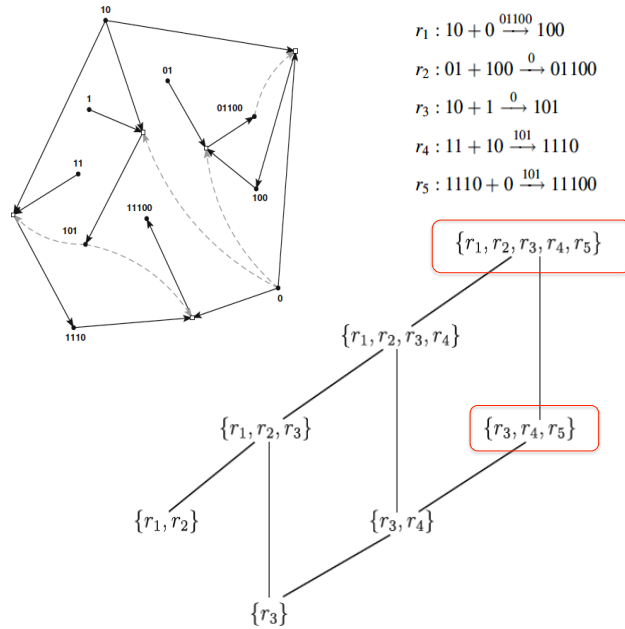

$$r_1 : 10 + 0 \xrightarrow{01100} 100$$
$$r_2 : 01 + 100 \xrightarrow{0} 01100$$
$$r_3 : 10 + 1 \xrightarrow{0} 101$$
$$r_4 : 11 + 10 \xrightarrow{101} 1110$$
$$r_5 : 1110 + 0 \xrightarrow{101} 11100$$

Figure 8: *Top:* The maxRAF (shown as a bipartite graph on the left, and a system of catalyzed reactions on the right) as found in an instance of the binary polymer model with $n = 5$ and $p = 0.0045$. *Bottom:* The poset of subRAFs, with the three closed subRAFs highlighted within rectangles. The closed RAF subset $\{r_3, r_4, r_5\}$ is also a CAF (and the maxCAF). Adapted from Hordijk et al. [30].

in this example, namely $\{r_1, r_2\}$ and $\{r_3\}$. Moreover, in this example, three of the subRAFs have the property of being *closed*, in the sense that each of these subRAFs contains every reaction r (that is possible in the binary polymer model) for which each reactant of r and at least one catalyst for r is either

produced by the subRAF itself or is present in the food set.

The maxRAF of any CRS is always closed. However, as in the example in Fig. 8, other (smaller) closed subRAFs often exist (by contrast, it can be shown that a CRS always has one closed CAF at most). Closed subRAFs tend to be more stable than non-closed subRAFs in dynamic simulations of CRSs, with the non-closed subRAFs forming briefly before either disappearing (due to diffusion) or enlarging to a closed subRAF. This dynamic picture of populations of subRAFs forming and changing within a CRS suggest a way in which autocatalytic sets might evolve.

In a recent study, Vasas et al. [64] concluded that autocatalytic sets *can* indeed be evolvable and that one of the necessary conditions for their evolvability is precisely the existence of multiple autocatalytic subsets within a given reaction network. The main idea is that different combinations of such autocatalytic subsets can co-exist (within "protocells"), giving rise to different "cell types" (or phenotypes). If these different phenotypes give rise to different growth rates (and other properties that are important for survival and reproduction), this can lead to selection and evolution. "Mutations" would then be the loss or gain of such autocatalytic subsets. Figure 9 shows another example of different RAF subsets existing within a maxRAF that appeared in an instance of the binary polymer model.
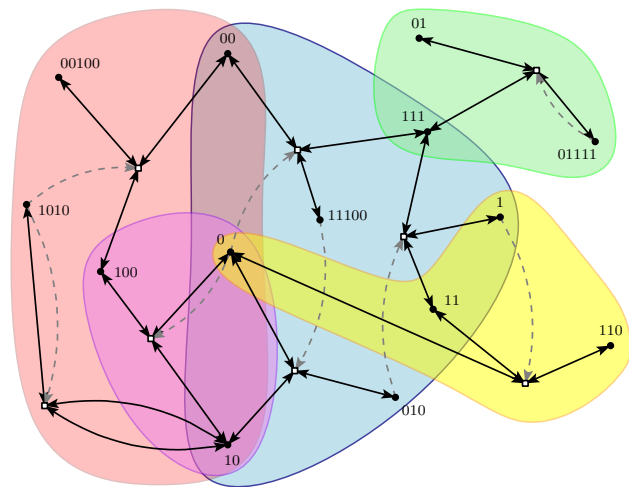

Figure 9: The maxRAF found in an instance of the binary polymer model with $n = 5$ and $p = 0.0045$, with the food set being monomers and dimers. The colored shapes indicate the various autocatalytic subsets residing within it; all of these are RAFs, except for the top right (green) subset, consisting of the single reaction $01 + 111 \leftrightarrow 01111$, which forms a co-RAF (as described in Section 2.1). Adapted from Hordijk and Steel [24].

It turns out that the requirement for "spontaneous" reactions is crucial for this potential evolvability [26]. Recall the simple example RAF shown in Figure 2 in the introduction. In this figure, neither of the two reactions making up this RAF set can proceed catalyzed when only food molecules are initially

present (the four reactants). At least one of these two reactions needs to happen uncatalyzed ("spontaneously") before the RAF set can actually be realized. Therefore, we also need to consider actual dynamics.

7. Dynamical and spatial aspects

Most of the results discussed so far are based on a "static" (or topological) reaction network analysis. However, as the requirement for spontaneous reactions indicates, the actual dynamics of these reaction networks is important too. Even if we know that autocatalytic (sub)sets exist within a given network (i.e., in a topological sense), it is not immediately clear how, or even if, they will emerge over time (i.e., in a dynamical sense).

Any chemical reaction can always happen uncatalyzed. However, this usually happens at a much lower rate compared to when it is catalyzed. Therefore, there may be a certain (stochastic) waiting time before a RAF set comes into existence when only food molecules are initially present. Consequently, in larger networks, different autocatalytic subsets will come into existence at different (random) times, due to the need for these spontaneous reactions.

These dynamical issues can be investigated using the Gillespie algorithm, a standard (and mathematically exact) stochastic method for simulating the dynamics of reaction networks [16, 17]. An example (in the form of a short movie) of different RAF subsets coming into existence at different times in the reaction network shown in Figure 9 is available at Hordijk [21]. The system starts with only food molecules, in which case only the yellow and purple RAF subsets can be realized. The other RAF subsets (red, blue, and green) each need at least one spontaneous reaction to happen, and therefore come into existence at different (random) times. In this particular example, first the blue one comes into existence and later on the red one. However, in another simulation run this might happen in the opposite order (due to the stochastic nature of the dynamics). In fact, if the rates of spontaneous reactions are low enough, only one of them, or even none, might come into existence. This way, different "protocells" could exist, by having different combinations of RAF subsets come into existence in different compartments.

Since this example is based on a simple polymer model, we are free to choose the reaction rates any way we want, and see how this influences the dynamics of the network. However, in other work we have used actual reaction rates as measured from chemical experiments. In that case, the dynamics of an experimental autocatalytic set of RNA molecules was simulated to study how it can form over time, and how environmental fluctuations can influence this process [32]. A similar simulation study using actual reaction rates is currently under way to study the dynamics of an experimental autocatalytic set of peptides.

The concept of protocells, mentioned in the previous paragraph, implies the presence of a boundary (e.g., a lipid membrane) in which the network dynamics are spatially contained. The notion of a boundary is not explicitly included in the formal RAF framework, but previously we have suggested how it can be incorporated implicitly by slightly generalizing the notion of catalysis [27]. To

see how, notice that a boundary ensures that molecules remain close together (rather than diffuse away), and therefore reactions between them will happen at a higher rate, while the boundary itself is not "used up" in these reactions.

In the stochastic simulation studies, boundaries (e.g. as semi-permeable membranes) can easily be included explicitly [65]. Also, we are currently performing dynamical simulations in a grid-like spatial environment, with a small amount of diffusion between neighboring cells. In these explicitly spatial situations, it seems that autocatalytic sets have an even higher chance of forming, and are also more stable, compared to a bulk solution. These results reflect similar conclusions from earlier simulation studies on replicator systems which were instrumental in pointing out the importance of spatial structure [5, 19].

Finally, localization of relevant molecules is not only achievable by membranes, but can also happen due to absorption on mineral surfaces, entrapment in eutectic phases in water-ice, or in naturally forming micro-pores such as in hydrothermal vents [47]. All of these cases could also be analyzed within the formal RAF framework, both in terms of (static) network analysis and (dynamic) stochastic simulations.

8. Relevance and generality

Many of the results reviewed in this article are based on applying RAF theory to the binary polymer model. However, the theory is not restricted to this abstract model of chemical reaction networks. For example, RAF theory has also been applied to formally study the experimental 16-member RNA autocatalytic set of Vaidya et al. [62]. Not only was it possible to reproduce many of the experimental results, but it also led to additional insights that would have been difficult to obtain from the experiments alone [25, 32].

Furthermore, RAF theory has been used to show that the metabolic network of *E. coli* contains an autocatalytic set of close to 1800 reactions, and to study the (modular) structure of this set in more detail [56]. Thus, autocatalytic sets are not just a theoretical concept, but they exist and can be studied in real chemical and biological reaction networks as well. Moreover, RAF theory captures the main results and properties of several different (computational) models that study the emergence and potential evolution of autocatalytic sets [28]. It thus provides a general and robust formal framework for performing such studies.

In fact, since RAF sets are defined in a graph-theoretical way, they are not restricted to chemical reaction networks only. In principle, the nodes in a "reaction graph" could represent any kind of entities and transformations between those entities. This way, the RAF framework can be extended and applied beyond chemistry and the origin of life. One such extension was suggested recently by viewing ecosystems as a network of mutually dependent autocatalytic sets (representing species), enabling each other's emergence and existence [6].

As another example, one could think of economic production functions as the equivalent of chemical reactions. Inputs such as wood and nails are transformed into outputs such as tables in an economic production function, just as

17

reactants are transformed into products in a chemical reaction (see Figure 10). Furthermore, some of those outputs, such as hammers, can act as "catalysts" in that they speed up the rate at which certain goods are produced, without being used up in that process. With this analogy in place, one could think of the economy as a whole as a catalytically closed and self-sustaining autocatalytic set [42, 20].
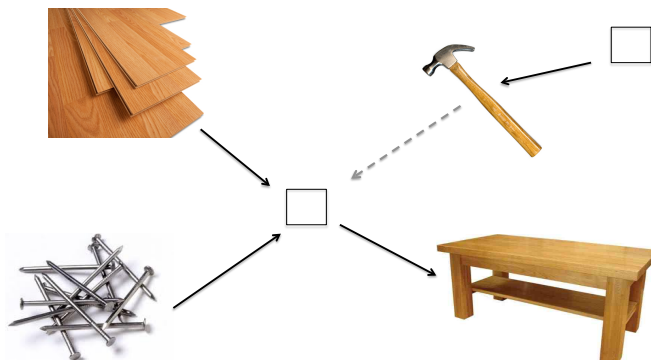

Figure 10: Considering economic production functions as the analog of chemical reactions, one could view the economy as a whole as a self-sustaining autocatalytic set.

9. Current and future work

This article has provided a general overview of the main ideas and the results of our mathematical and computational studies of autocatalytic sets, formalized as RAF theory. There are still other aspects of interest, such as using an infinite maximum molecule length in the binary polymer model [58], or using more realistic ways to assign catalysts rather than assigning them purely randomly [29, 33, 31], but we do not have the space here to describe them all in detail. And, of course, much work still remains to be done. Some of the things we are currently working on or hope to work on in the near future include:

- Studying additional real chemical and biological networks using the formal RAF framework, such as an experimental peptide autocatalytic set [1], or the metabolic network of certain prokaryotes that are believed to be much closer (in evolutionary terms) to the last universal common ancestor (LUCA) than *E. coli*;

- Using actual molecular structure, such as folded RNA sequences, to determine catalysis in models of chemical reaction networks;

- Simulating the dynamics of autocatalytic sets within growing and dividing compartments with a (semi-)permeable boundary;

18

- Formalizing and applying the notion of autocatalytic sets beyond chemistry, such as in ecosystems or even the economy.

The road ahead may not be easy, but has many exciting prospects and promising views. We hope that with the work done so far, as described here, we have laid down a solid (and formal) foundation for a continued journey ahead.

Acknowledgments

References

[1] G. Ashkenasy, R. Jegasia, M. Yadav, and M. R. Ghadiri. Design of a directed molecular network. *PNAS*, 101(30):10872–10877, 2004.

[2] R. J. Bagley and J. D. Farmer. Spontaneous emergence of a metabolism. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pages 93–140. Addison-Wesley, 1991.

[3] R. J. Bagley, J. D. Farmer, and W. Fontana. Evolution of a metabolism. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pages 141–158. Addison-Wesley, 1991.

[4] S. A. Benner, H. J. Kim, and Z. Yang. Setting the stage: the history, chemistry, and geobiology behind RNA. *Cold Spring Harbour Perspectives in Biology*, 4:a003541, 2012.

[5] M. A. Boerlijst and P. Hogeweg. Spiral wave structure in pre-biotic evolution: Hypercycles stable against parasites. *Physica D*, 48:17–28, 1991.

[6] R. Cazzolla Gatti, W. Hordijk, and S. Kauffman. Biodiversity is autocatalytic. *Ecological Modelling*, 2016. In press.

[7] C. Darwin. *On the Origin of Species*. John Murray, 1859.

[8] P. Dittrich and P. Speroni di Fenizio. Chemical organization theory. *Bulletin of Mathematical Biology*, 69(4):1199–1231, 2007.

[9] M. Eigen and P. Schuster. *The Hypercycle*. Springer-Verlag, 1979.

[10] P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.

[11] P. Erdős and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

[12] J. D. Farmer, S. A. Kauffman, and N. H. Packard. Autocatalytic replication of polymers. *Physica D*, 22:50–67, 1986.

[13] 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*, 2:2, 2011.

[14] A. Filisetti, M. Villani, C. Damiani, A. Graudenzi, A. Roli, W. Hordijk, and R. Serra. On RAF sets and autocatalytic cycles in random reaction networks. *Communications in Computer and Information Science*, 445: 113–126, 2014.

[15] W. Gilbert. The RNA world. *Nature*, 319:618, 1986.

[16] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434, 1976.

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

[18] P. G. Higgs and N. Lehman. The RNA world: molecular cooperation at the origins of life. *Nature Review Genetics*, 16:7–17, 2015.

[19] P. Hogeweg and N. Takeuchi. Multilevel selection in models of prebiotic evolution: compartments and spatial self-organization. *Origins of Life and Evolution of Biospheres*, 33:375–403, 2003.

[20] W. Hordijk. Autocatalytic sets: From the origin of life to the economy. *BioScience*, 63(11):877–881, 2013.

[21] W. Hordijk. Autocatalytic sets arising. http://www.WorldWideWanderings.net/blog/?p=561, 2016.

[22] W. Hordijk and M. Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227(4):451–461, 2004.

[23] W. Hordijk and M. Steel. Predicting template-based catalysis rates in a simple catalytic reaction model. *Journal of Theoretical Biology*, 295:132–138, 2012.

[24] W. Hordijk and M. Steel. Autocatalytic sets extended: Dynamics, inhibition, and a generalization. *Journal of Systems Chemistry*, 3:5, 2012.

[25] W. Hordijk and M. Steel. A formal model of autocatalytic sets emerging in an RNA replicator system. *Journal of Systems Chemistry*, 4:3, 2013.

[26] W. Hordijk and M. Steel. Conditions for evolvability of autocatalytic sets: A formal example and analysis. *Origins of Life and Evolution of Biospheres*, 44(2):111–124, 2014.

[27] W. Hordijk and M. Steel. Autocatalytic sets and boundaries. *Journal of Systems Chemistry*, 6:1, 2015.

[28] W. Hordijk and M. Steel. Autocatalytic sets in polymer networks with variable catalysis distributions. *Journal of Mathematical Chemistry*, 54 (10):1997–2021, 2016.

[29] W. Hordijk, S. A. Kauffman, and M. Steel. Required levels of catalysis for emergence of autocatalytic sets in models of chemical reaction systems. *International Journal of Molecular Sciences*, 12(5):3085–3101, 2011.

[30] W. Hordijk, M. Steel, and S. Kauffman. The structure of autocatalytic sets: Evolvability, enablement, and emergence. *Acta Biotheoretica*, 60(4): 379–392, 2012.

[31] W. Hordijk, L. Hasenclever, J. Gao, D. Mincheva, and J. Hein. An investigation into irreducible autocatalytic sets and power law distributed catalysis. *Natural Computing*, 13(3):287–296, 2014.

[32] W. Hordijk, N. Vaidya, and N. Lehman. Serial transfer can aid the evolution of autocatalytic sets. *Journal of Systems Chemistry*, 5:4, 2014.

[33] W. Hordijk, P. R. Wills, and M. Steel. Autocatalytic sets and biological specificity. *Bulletin of Mathematical Biology*, 76(1):201–224, 2014.

[34] W. Hordijk, J. I. Smith, and M. Steel. Algorithms for detecting and analysing autocatalytic sets. *Algorithms for Molecular Biology*, 10:15, 2015.

[35] D. P. Horning and G. F. Joyce. Amplification of RNA by an RNA polymerase ribozyme. *PNAS*, 113:9786–9791, 2016.

[36] C. J. Jeffery. Moonlighting proteins. *Trends in Biochemical Sciences*, 24 (1):8–11, 1999.

[37] C. J. Jeffery. Moonlighting proteins: Old proteins learning new tricks. *Trends in Genetics*, 19(8):415–417, 2003.

[38] G. F. Joyce. The antiquity of RNA-based evolution. *Nature*, 418:214–221, 2002.

[39] S. A. Kauffman. Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. *Journal of Cybernetics*, 1(1): 71–96, 1971.

[40] S. A. Kauffman. Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119:1–24, 1986.

[41] S. A. Kauffman. *The Origins of Order*. Oxford University Press, 1993.

[42] S. A. Kauffman. Economics and the collectively autocatalytic structure of the real economy. http://www.npr.org/blogs/13.7/2011/11/21/142594308/economics-and-the-collectively-autoca 2011.

[43] D.-E. Kim and G. F. Joyce. Cross-catalytic replication of an RNA ligase ribozyme. *Chemistry & Biology*, 11:1505–1512, 2004.

[44] S. Lifson. On the crucial stages in the origin of animate matter. *Journal of Molecular Evolution*, 44:1–8, 1997.

[45] T. A. Lincoln and G. E. Joyce. Self-sustained replication of an RNA enzyme. *Science*, 323:1229–1232, 2009.

[46] J. Maynard Smith. Hypercycles and the origin of life. *Nature*, 280:445–446, 1979.

[47] P.-A. Monnard. Taming prebiotic chemistry: The role of heterogeneous and interfacial catalysis in the emergence of a prebiotic catalytic/information polymer system. *Life*, 6:40, 2016.

[48] E. Mossel and M. Steel. Random biochemical networks: The probability of self-sustaining autocatalysis. *Journal of Theoretical Biology*, 233(3):327–336, 2005.

[49] P. Nghe, W. Hordijk, S. A. Kauffman, S. I. Walker, F. J. Schmidt, H. Kemble, J. A. M. Yeates, and N. Lehman. Prebiotic network evolution: six key parameters. *Molecular BioSystems*, 11:3206–3217, 2015.

[50] L. Pasteur. Des générations spontanées. *Revues des Cours Scientifiques*, 1 (21):257–265, 1864.

[51] V. Patzke and G. von Kiedrowski. Self replicating systems. *Arkivoc*, 2007 (5):293–310, 2007.

[52] M. W. Powner, B. Gerland, and J. D. Sutherland. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, 459:239–242, 2009.

[53] R. Rosen. *Life Itself*. Columbia University Press, 1991.

[54] D. Sievers and G. von Kiedrowski. Self-replication of complementary nucleotide-based oligomers. *Nature*, 369:221–224, 1994.

[55] J. Smith, M. Steel, and W. Hordijk. Autocatalytic sets in a partitioned biochemical network. *Journal of Systems Chemistry*, 5:2, 2014.

[56] F. L. Sousa, W. Hordijk, M. Steel, and W. F. Martin. Autocatalytic sets in *E. coli* metabolism. *Journal of Systems Chemistry*, 6:4, 2015.

[57] M. Steel. The emergence of a self-catalysing structure in abstract origin-of-life models. *Applied Mathematics Letters*, 3:91–95, 2000.

[58] M. Steel. Self-sustaining autocatalytic networks within open-ended reaction systems. *Journal of Mathematical Chemistry*, 53(8):1687–1701, 2015.

[59] M. Steel, W. Hordijk, and J. Smith. Minimal autocatalytic networks. *Journal of Theoretical Biology*, 332:96–107, 2013.

[60] E. Szathmáry. On the propagation of a conceptual error concerning hypercycles and cooperation. *Journal of Systems Chemistry*, 4:1, 2013.

[61] J. W. Szostak. The eightfold path to non-enzymatic RNA replication. *Journal of Systems Chemistry*, 3:2, 2012.

[62] N. Vaidya, M. L. Manapat, I. A. Chen, R. Xulvi-Brunet, E. J. Hayden, and N. Lehman. Spontaneous network formation among cooperative RNA replicators. *Nature*, 491:72–77, 2012.

[63] V. Vasas, E. Szathmáry, and M. Santos. Lack of evolvability in self-sustaining autocatalytic networks constraints metabolism-first scenarios for the origin of life. *PNAS*, 107(4):1470–1475, 2010.

[64] V. Vasas, C. Fernando, M. Santos, S. Kauffman, and E. Sathmáry. Evolution before genes. *Biology Direct*, 7:1, 2012.

[65] M. Villani, A. Filisetti, A. Graudenzi, C. Damiani, T. Carletti, and R. Serra. Growth and division in a dynamic protocell model. *Life*, 4:837–864, 2014.

[66] R. Wolfenden and M. J. Snider. The depth of chemical time and the power of enzymes as catalysts. *Accounts of Chemical Research*, 34(12):938–945, 2001.