Autocatalytic Sets Arising in a Combinatorial Model of Chemical Evolution

Wim Hordijk [1,]*, **Mike Steel** [2] and **Stuart Kauffman** [3]

1   SmartAnalytiX, 1170 Vienna, Austria
2   Biomathematics Research Centre, University of Canterbury, Christchurch 8041, New Zealand
3   Institute for Systems Biology, Seattle, WA 98109, USA
*   Correspondence: wim@worldwidewanderings.net

Abstract: The idea that chemical evolution led to the origin of life is not new, but still leaves open the question of how exactly it could have led to a coherent and self-reproducing collective of molecules. One possible answer to this question was proposed in the form of the emergence of an autocatalytic set: a collection of molecules that mutually catalyze each other's formation and that is self-sustaining given some basic "food" source. Building on previous work, here we investigate in more detail when and how autocatalytic sets can arise in a simple model of chemical evolution based on the idea of combinatorial innovation with random catalysis assignments. We derive theoretical results, and compare them with computer simulations. These results could suggest a possible step towards the (or an) origin of life.

1. Introduction

The idea that chemical evolution led to the origin of life was proposed independently by Oparin [1] and Haldane [2]. It was later shown to be plausible through the experiments of Miller [3]. Recently, different variants of these experiments were repeated, and analyzed with state-of-the-art molecular analysis technology [4,5]. This revealed the presence of thousands of molecular species, including many classes of organic and catalytic ones. However, it still leaves open the question of how such spontaneous chemical evolution can give rise to a coherent, self-reproducing collective of molecules.

One possible answer to this question was proposed in the form of the emergence of autocatalytic sets [6,7]. Informally, an autocatalytic set is a chemical reaction network in which the molecules mutually catalyze each other's formation, and which is self-sustaining given a basic food source. That such autocatalytic sets can indeed form spontaneously was already shown early on through computer simulations [8–10]. Later on, they were also successfully constructed with real molecules in the lab [11–14], and shown to exist in the metabolic networks of prokaryotes [15–17]. The notion of autocatalytic sets was formalized and studied extensively as reflexively autocatalytic and food-generated (RAF) sets (see, for example, [18,19] and references therein).

Recently, a simple model of combinatorial innovation, referred to as the *theory of the adjacent possible* (TAP), was studied formally [20]. In general, TAP states that evolving systems create their own future possibilities in an ever-increasing "adjacent possible". In particular, the number of "things" that can come into existence (or be created) next increases as a combinatorial function of what is currently in existence. As such, this model can also be interpreted in the context of chemical evolution, where the "things" are molecular species. The more molecular species that are currently in existence, the more potential new species that can come into existence next through (spontaneous) chemical reactions between arbitrary combinations of currently existing molecules. An initial investigation of the combination of the TAP and RAF models was presented in the context of technological

evolution [21]. Here, we present results of a further investigation into this model combination, with both theoretical and computer simulation results, and more specifically in the context of chemical evolution and how it could lead to the formation of mutually catalytic and self-reproducing collectives of molecules as a possible step towards the (or an) origin of life.

2. Methods

2.1. TAP

The TAP model is based on the idea of combinatorial innovation [20]. At its core is the following equation from [22]:

$$M_{t+1} = M_t + \sum_{i=1}^{M_t} \alpha_i \binom{M_t}{i}, \tag{1}$$

where M_t is the number of different "things" at time t, and α_i is a decreasing sequence of probabilities (i.e., real numbers between 0.0 and 1.0). Interpreted in the context of chemical evolution, new molecular species are produced through chemical reactions with arbitrary combinations of already existing species as reactants. More specifically, at each time step t, each possible combination of i existing molecular species has a small probability α_i of chemically reacting, producing a new species. In other words, the probabilities α_i could be considered some sort of reaction propensities.

Note, though, that the above equation represents a deterministic version that does not guarantee M_t to be an integer value, and it only serves to convey the general idea behind the model. Instead, we use a stochastic implementation that does guarantee integer values, as in [20]. Although our mathematical results do not require it, we will assume for the simulations and the following algorithm that $\alpha_i = \alpha^i$ (i.e., α to the power i).

In addition, for each newly created molecular species x and each of the existing molecule types y that were not present at time 0, x can catalyze the formation of y with a fixed probability p. Similarly, for each chemical reaction r that produces a new species, and each of the existing molecule types y that were not present at time 0, y can catalyze r, also with probability p. These random catalysis assignments are assumed to be independent across all pairs of molecules and reactions.

Our implementation of this model is described in Algorithm 1.

Several remarks should be made about this algorithmic description. First, an upper limit K on the possible number of reactants is set for numerical and computational reasons. It was already shown earlier that this does not significantly affect the overall behavior of the TAP model [20], and chemically it is a plausible constraint as well.

Next, the algorithm stops when *exactly* **M** molecular species have been produced (due to the *if*-statement at line 16). Previously, the algorithm was allowed to finish the time step in which **M** items are reached, in which case the final amount of items is generally larger than **M** [21]. However, for a more accurate comparison with theoretical results, here we terminate the algorithm immediately after the **M**$^{\text{th}}$ molecular species has been produced, thus not finishing the rest of the time step in which that happens.

Finally, in the *for*-loop at line 12, a new molecular species x is considered twice to catalyze its own production (when $y = M_t$). However, the loop is stated this way to keep it more concise, rather than having an additional *if*-statement, or two separate *for*-loops. Of course, in an actual implementation this double consideration can be easily excluded, and as long as M_t is large enough, it is negligible in the theoretical derivations below.

Algorithm 1 TAP with catalysis.

Require: $M_0, \alpha, K, \mathbf{M}, p$

1:   $t \leftarrow 0$
2:   Create M_0 initial molecular species labeled $1, \ldots, M_0$
3: **while** $M_t < \mathbf{M}$ **do**
4:     $t \leftarrow t + 1; M_t \leftarrow M_{t-1}$
5:     **for** $i = 1, \ldots, K$ **do**
6:         $s_i \leftarrow \alpha^i \times \binom{M_{t-1}}{i}$
7:         $r_i \leftarrow \text{Poisson}(s_i)$
8:         **for** $j = 1, \ldots, r_i$ **do**
9:             Create a new species x labeled $M_t + 1$
10:            Select i random reactants for the production of x from $1, \ldots, M_{t-1}$
11:            $M_t \leftarrow M_t + 1$
12:            **for** $y = M_0 + 1, \ldots, M_t$ **do**
13:               With probability p assign x as catalyst to the reaction that produced y
14:               With probability p assign y as catalyst to the reaction that produced x
15:            **end for**
16:            **if** $M_t = \mathbf{M}$ **then**
17:               $i \leftarrow K + 1; j \leftarrow r_i + 1$
18:            **end if**
19:         **end for**
20:     **end for**
21: **end while**

Over time, the M_t existing molecular species, and the particular chemical reactions that produced them, thus form a growing chemical reaction network, with a "food set" consisting of the M_0 initial species, as in the example in Figure 1. In addition, the molecules catalyze each other's formation according to the catalysis probability p. Note that each reaction can have no, one, or multiple catalysts, depending on these random catalysis assignments. Similarly, each molecule type may catalyze no, one, or multiple reactions. Figure 1 shows a simple example of a chemical reaction network that resulted from the TAP model, with random catalysis assigned as indicated by the dashed arrows.
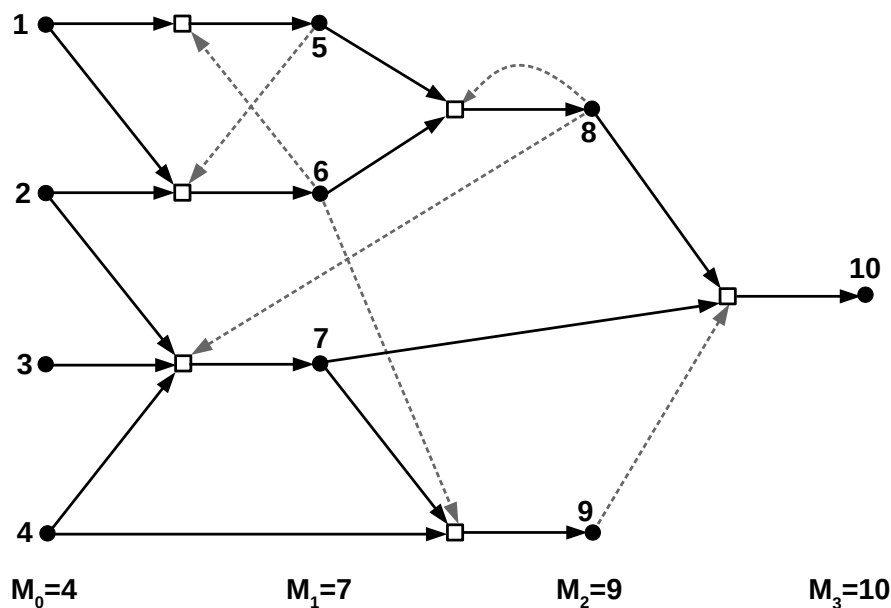

Figure 1. An example chemical reaction network resulting from the TAP model, with random catalysis assignments (dashed arrows). Modified from [21].

As before [21], the parameters M_0, α, and K are fixed at the values $M_0 = 10$, $\alpha = 0.01$, and $K = 4$.

2.2. RAF

One could now ask what the probability is that in such a growing chemical reaction network, at some point a subset of molecule types exists in which the molecules mutually catalyze each other's production, and which is sustainable on the given food set. Such a subset is known as a reflexively autocatalytic and food-generated (RAF) set [18]. More formally, an RAF set is a set $\mathcal{R}$ of chemical reactions and the molecule types involved in them such that:

1. Each reaction in $\mathcal{R}$ is catalyzed by at least one of the molecules involved in $\mathcal{R}$.
2. Each molecule type involved in $\mathcal{R}$ can be created from the food set through a sequence of reactions from $\mathcal{R}$ itself.

An efficient computer algorithm exists to find such RAF sets in arbitrary chemical reaction networks, or determine that no such subset is present. This algorithm actually finds the (unique) *maximal* RAF (maxRAF), i.e., the union of all possible RAFs within a given network. Repeated application of the algorithm can then also identify smaller RAF subsets within the maxRAF, including minimal ones.

In fact, the entire reaction network in Figure 1 forms a maxRAF, but it contains several smaller RAF subsets (e.g., the two reactions forming the molecule types $\{5, 6\}$ are an RAF, and so too are the four reactions forming $\{5, 6, 7, 8\}$).

2.3. TAP and RAF

An initial investigation into the existence of RAF sets in the TAP model with catalysis was presented recently, but in the context of technological evolution [21]. Here, we provide a more detailed study, and more specifically in the context of chemical evolution. First, we derive theoretical expressions for the probability of RAFs existing in instances of the TAP model. We then compare these with results from computer simulations, using an implementation of the TAP model as presented in Algorithm 1, and applying the RAF algorithm to large sets of random instances of the TAP model.

3. Results

Consider an instance of the TAP model, described by $(\mathcal{M}_t, \mathcal{R}_t)$ (for $t = 0, 1, 2, \ldots$) where $\mathcal{M}_t$ is a set of molecular species generated up to time t and $\mathcal{R}_t$ is the set of all reactions involved in generating $\mathcal{M}_t$ starting from $\mathcal{M}_0$. We let $M_t = |\mathcal{M}_t|$ and $R_t = |\mathcal{R}_t|$ denote the number of molecular species in $\mathcal{M}_t$ and reactions in $\mathcal{R}_t$, respectively. Note that these families of sets (and their sizes) are random variables for each $t \geq 1$. We assume throughout this section that $M_0 \geq 1$, and $\alpha_i \neq 0$ for at least one value of $i \leq M_0$. We also explicitly assume in this section that for each pair (x, r) where $x \in \mathcal{M}_t \setminus \mathcal{M}_0$ and $r \in \mathcal{R}_t$, x catalyzes r with probability p, and that these catalysis events are stochastically independent.

Throughout this section we do not specifically require that $\alpha_i = \alpha^i$, nor do we place any upper bound (such as **M**) on M_t unless otherwise stated, or any bound (such as K) on the number of reactants of a reaction.

Since every reaction in the TAP model creates exactly one new product, we have:

$$R_t = M_t - M_0. \tag{2}$$

Lemma 1. *In the TAP model, M_t is nondecreasing, and with probability 1, $M_t \to \infty$ as t grows.*

Proof. M_t is a Markovian random walk on the positive integers, with $M_{t+1} \geq M_t$ for all t, and the probability of the event $M_{t+1} - M_t \geq 1$ is uniformly bounded away from 0 for all values of t. By a standard probability argument, for any positive integer k, the event E_k that $M_t \leq k$ holds for all t has probability 0. Thus, $\mathbb{P}(\cup_{k \geq 1} E_k) = 0$ which ensures that $M_t \to \infty$ with probability 1. $\square$

Remark 1. *If an upper bound* **M** *is imposed on M_t (so that the process terminates when $M_t \geq$ **M***)
then Lemma 1 implies that M_t is certain to eventually hit* **M**. *Note also that the version of the TAP
model studied theoretically in [20] takes place in continuous (rather than discrete) time, in which
case Lemma 1 has a sharper statement: Provided that $\alpha_1 > 0$ and $\alpha_i > 0$ for at least one other value
of i, then with probability 1, M_t tends to infinity in* finite *time. Here, we are modeling a system in
discrete time, and so this "explosion in finite time" phenomenon does not arise. Nevertheless, in
our simulation results, where we explicitly stop the process when* **M** *molecular species have been
produced, a sudden and rapid increase still occurs.*

Next, consider the probability that the entire collection of reactions involved in generating $\mathcal{M}_t$ (i.e., $\mathcal{R}_t$) is an RAF. As the following lemma shows, this probability depends only on $\mathcal{M}_t$ only through the size of this set (i.e., M_t), and so we denote the probability by $P_{\text{all}}(M_t)$.

Lemma 2.
$$P_{\text{all}}(M_t) = \left(1 - (1-p)^{M_t - M_0}\right)^{M_t - M_0}. \tag{3}$$

Proof. $\mathcal{R}_t$ is *F*-generated, and so it forms an RAF precisely if each reaction in $\mathcal{R}_t$ is catalyzed by at least one molecule type in $\mathcal{M}_t \setminus \mathcal{M}_0$. By the independence assumption concerning catalysis assignments, the probability that any given reaction $r \in \mathcal{R}_t$ is catalyzed by at least one molecular species in $\mathcal{M}_t \setminus \mathcal{M}_0$ is $1 - (1-p)^{M_t - M_0}$, and since there are R_t reactions, the probability that all reactions in $\mathcal{R}_t$ are catalyzed is $\left(1 - (1-p)^{M_t - M_0}\right)^{R_t}$ which equals $\left(1 - (1-p)^{M_t - M_0}\right)^{M_t - M_0}$ by Equation (2). $\square$

We can now state our main theorem.

Theorem 1. *For the TAP + catalysis model, the following hold.*

(a) *For any value of $p > 0$, $P_{\text{all}}(M_t) = 1 - o(1)$, where $o(1)$ is a term that tends to zero as $M_t \to \infty$.*

(b) *Suppose that $M_t = m$ at some fixed time $t > 0$. Then the following hold:*

  (i) *If $p = \frac{\ln(xm)}{m}$ for $xm > 1$, then*

$$P_{\text{all}}(M_t) = \exp\left(-\frac{1}{x}\right) + o(1),$$

  where $o(1)$ is a term that converges to 0 as $m \to \infty$.

  (ii) *The expected number of reactions that each molecule type (not in $\mathcal{M}_0$) catalyzes (i.e., $f = pR_t = p(M_t - M_0)$) in order for $P_{\text{all}}(M_t)$ to equal θ is given by:*

$$f = \ln(m) + \ln\left(\frac{1}{\ln(1/\theta)}\right) + o(1),$$

  where $o(1) \to 0$ as $m \to \infty$. In particular, for each such value of θ, f grows logarithmically with m.

Proof.
Part (a): Let $M'_t = M_t - M_0$, and let $q = (1-p)^{M'_t}$. By the inequality $1 - x \leq e^{-x}$ for $x > 0$, we obtain $q \leq e^{-pM'_t}$, and thus, by Lemma 2, we obtain:

$$P_{\text{all}}(M_t) = (1-q)^{M'_t} \geq 1 - M'_t q,$$

(since $(1-x)^n \geq 1 - nx$ for $0 < x < 1$). Thus, $P_{\text{all}}(M'_t) \geq 1 - M'_t e^{-pM'_t} \to 1$ as $M_t \to \infty$.

Part (b-i): Conditional on $M_t = m$, and setting $p = \frac{\ln(xm)}{m}$, Lemma 2 gives:

$$\ln(P_{\text{all}}(M_t)) = (m - M_0) \cdot \ln\left[1 - \left(1 - \frac{\ln(xm)}{m}\right)^{m-M_0}\right]$$

$$\sim m \cdot \ln\left(1 - e^{-\ln mx}\right) = m \cdot \ln\left(1 - \frac{1}{mx}\right) \sim -\frac{1}{x},$$

where $\sim$ refers to asymptotic identity as m grows. Exponentiating gives $P_{\text{all}}(M_t) = \exp\left(-\frac{1}{x}\right) + o(1)$, as required.

Part (b-ii): Conditional on $M_t = m$, and setting $p = \frac{\ln(xm)}{m}$ for a value $x > 0$ (to be determined), gives:

$$f = pR_t = p(M_t - M_0) = p(m - M_0) = \ln(m) + \ln(x) - pM_0. \tag{4}$$

From Part (b), and again conditional on $M_t = m$, the equation $P_{\text{all}}(M_t) = \theta$ gives $x = \frac{1}{\ln(1/\theta)} + o(1)$ where $o(1)$ tends to 0 as m grows. Noting also that $pM_0 = o(1)$, the result now follows from Equation (4). $\square$

To see how fast $P_{\text{all}}(M_t)$ converges to 1, Figure 2 shows the theoretical probability of Equation (3) for an all-M_t RAF (solid line) against M_t for a catalysis probability $p = 0.005$. The open circles represent results from the TAP model simulations. The dashed line shows simulation results for any-sized RAF (i.e., an RAF consisting of any number of reactions).
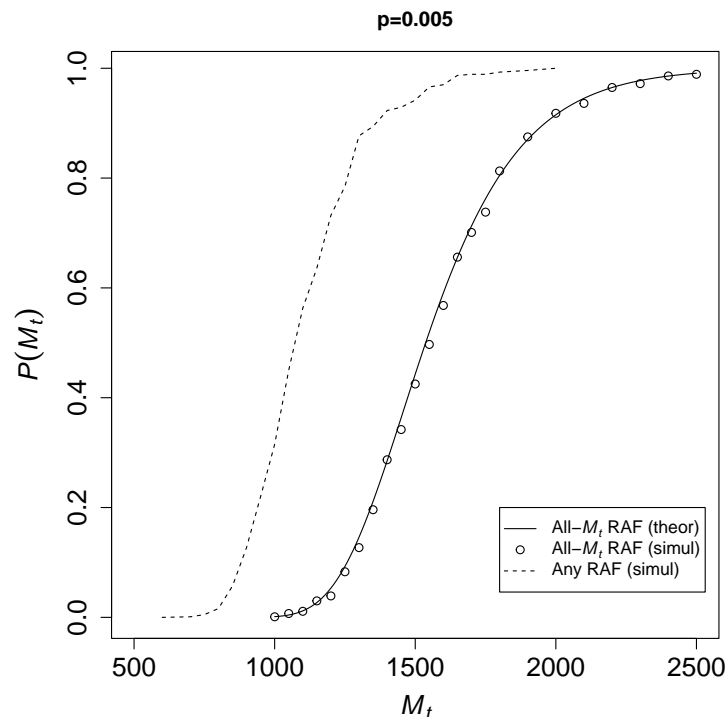

Figure 2. The probability $P(M_t)$ of RAFs against M_t for a catalysis probability $p = 0.005$.

Given a fixed probability of catalysis p, it is clear that once the total number of molecular species M_t becomes large enough, there is a sharp transition from RAF sets not existing at all to them existing in almost every instance of the model. Of course, RAFs of any size (dotted line) already occur at smaller values of M_t than all-M_t RAFs, but theoretically it is easier to deal with all-M_t RAFs, as they are always automatically food-generated. Thus, the theoretical expression forms an upper bound on the actual probabilities $P(M_t)$.

Another way to consider these probabilities is to fix the number of molecular species M_t and then see what the required level of catalysis $f = pM_t$ is to obtain RAF sets with high

probability. This level of catalysis indicates the average number of reactions catalyzed per molecule type. To see how this increases with increasing M_t, Figure 3 shows the theoretical probability (solid lines) of an all-M_t RAF against pM_t for different values of M_t. The dots are values obtained from computer simulations of the TAP model, to again compare with the theoretical results. As expected, the curves move slowly to the right for larger values of M_t, but the distance between each next pair of adjacent curves seems to be decreasing.
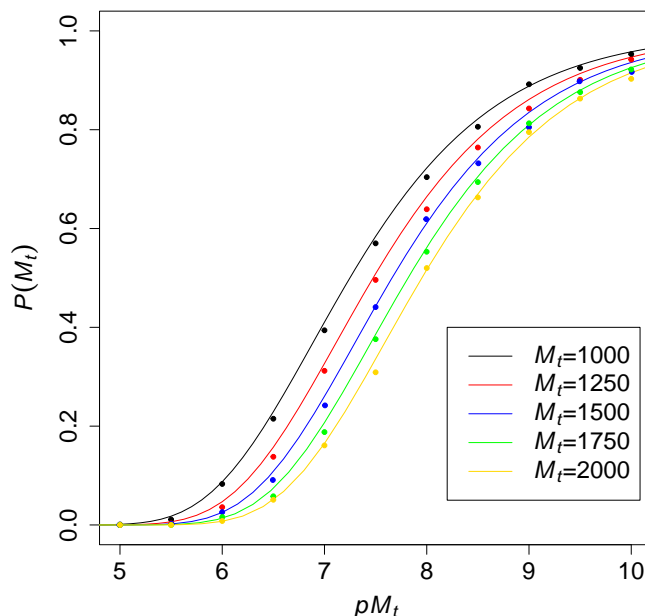
Figure 3. The probability $P(M_t)$ of all-M_t RAFs against pM_t for different values of M_t.

Taking the "transition point" to a high probability of RAFs to be at $P_{\text{all}}(M_t) = 0.5$, Theorem 1(b-ii) predicts that $f = pM_t$ should be close to $\ln(M_t) + \ln(1/\ln(2)) \approx \ln(M_t) + 0.367$. Figure 4 shows this function for a range of values of M_t, with the open circles representing results from computer simulations (interpolated from the simulation data shown in Figure 3). These simulation results closely fit with the theoretically predicted logarithmic curve.
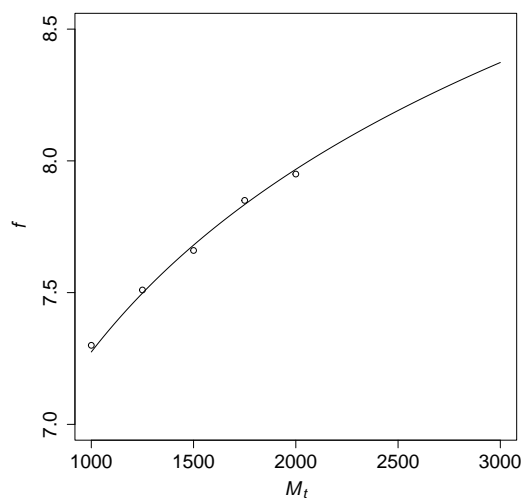

Figure 4. The required level of catalysis $f = pM_t$ to obtain all-M_t RAFs with a probability 0.5, for a range of M_t values.

Finally, a comparison is made between the probabilities of any-sized RAFs previously obtained from simulations of the TAP model [21], and earlier results from a related model

known as the binary polymer model [23]. In this related model, smaller polymers can ligate into larger and larger ones, and larger polymers can cleave into smaller and smaller ones. This model has been investigated extensively in the context of RAF sets in the past [18,19].

Figure 5 shows the probability P of an RAF against the level of catalysis $f = pR$ for various versions of the binary polymer model that use different ways of assigning catalysis. The red curves are the standard (uniform) catalysis distribution, the blue curves are a power law catalysis distribution, and the green lines are a sparse catalysis distribution [23]. These results were obtained from computer simulations. The black curves are the theoretically calculated probabilities for an all-or-nothing catalysis distribution. Solid lines are maximum polymer length $n = 10$, while dashed lines are $n = 16$. The thick gray line shows simulation results for the TAP model, with an average number of molecular species of $M_t = 1250$.
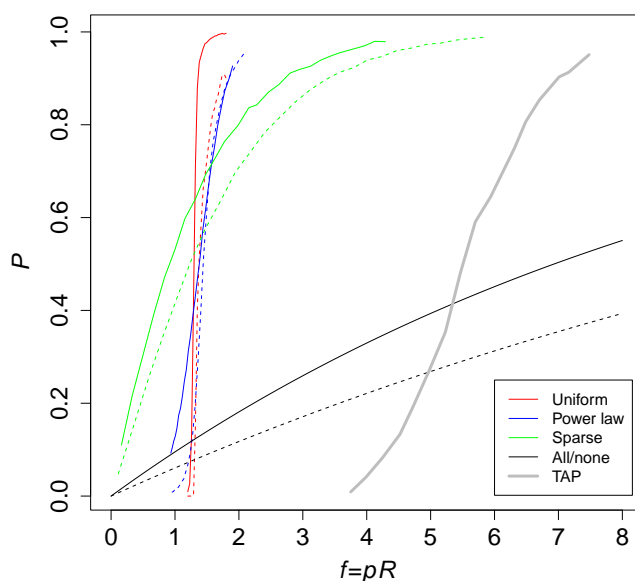

Figure 5. The probability P of RAFs in different versions of the binary polymer model (solid lines for $n = 10$ and dashed lines for $n = 16$) and in the TAP model ($M_t = 1250$). Adapted from Hordijk and Steel [23].

Clearly, the required level of catalysis to cause RAF sets to arise is higher in the TAP model (five to seven reactions catalyzed per molecule type, on average) than in the binary polymer model (one to two). However, this can be explained by the fact that in the TAP model there is always only one reaction that produces a given molecular species, whereas in the binary polymer model there are multiple reactions that can produce a given polymer. In other words, there is a large amount of redundancy in the reaction networks resulting from the binary polymer model, allowing for a lower level of catalysis (given that only one or two of the multiple reactions that produce a given polymer need to be catalyzed).

Note also that we restricted the chemical reactions in the TAP model to only generate one product. In general, reactions can produce more than one product, in which case there could be significantly more molecular species than reactions. It is therefore expected that in such a more general model version the required level of catalysis f will be lower than the five to seven suggested by Figure 5.

4. Conclusions

We reinterpreted a simple model of combinatorial innovation known as TAP (theory of the adjacent possible) in the context of chemical evolution and autocatalytic sets. We then derived theoretical expressions for the probabilities of such autocatalytic sets arising in instances of the TAP model. These theoretical predictions were verified with results from computer simulations.

These results show that autocatalytic sets do indeed have a high probability of arising in instances of the TAP model, given a large enough number of molecular species and/or level of catalysis. Of course this is still a very general model, as it is assumed that any molecular species can chemically react with any other, and catalysis is assigned randomly. However, previous work on a related model, known as the binary polymer model, showed that more realistic assumptions can be easily incorporated in such a model, and do not change the overall results very much, at least not qualitatively. Moreover, the quantitative changes can often be predicted from the more general basic model version [18].

Generally, the level of catalysis (i.e., the average number of reactions catalyzed per molecule type) needs to be somewhat higher in the TAP model than in the binary polymer model. However, this can be explained by the redundancy present in reaction networks resulting from the binary polymer model. One could easily imagine a version of the TAP model where certain molecular species can also be produced by multiple reactions.

In conclusion, the results presented here may suggest a possible step towards the (or an) origin of life, where self-sustaining and reproducing autocatalytic sets arise during a process of chemical evolution. In fact, Wollrab et al. [4] conclude from their "Miller-type" chemical evolution experiments that "organic catalysts that appear in the broth may well lead to the production of molecular species that would normally not be favored under the conditions in the reactor, further enhancing the molecular richness". If even just some of those species happen to form a closed loop, mutually catalyzing each other, autocatalytic sets would indeed arise spontaneously.

References

1. Oparin, A.I. *The Origin of Life*; Moscow Worker Publishing: Moscow, Russia, 1924. (In Russian)
2. Haldane, J.B.S. The origin of life. *Ration. Annu.* **1929**, 148, 3–10.
3. Miller, S.L. Production of amino acids under possible primitive earth conditions. *Science* **1953**, *117*, 528–529. [CrossRef] [PubMed]
4. Wollrab, E.; Scherer, S.; Aubriet, F.; Carré, V.; Carlomagno, T.; Codutti, L.; Ott, A. Chemical analysis of a "Miller-type" complex prebiotic broth Part I: Chemical diversity, oxygen and nitrogen based polymers. *Orig. Life Evol. Biosph.* **2015**, *46*, 149–169. [CrossRef] [PubMed]
5. Scherer, S.; Wollrab, E.; Codutti, L.; Carlomagno, T.; Gomes da Costa, S.; Volkmer, A.; Bronja, A.; Schmitz, O.J.; Ott, A. Chemical analysis of a "Miller-type" complex prebiotic broth Part II: Gas, oil, water and the oil/water-interface. *Orig. Life Evol. Biosph.* **2017**, *47*, 381–403. [CrossRef] [PubMed]
6. Kauffman, S.A. Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. *J. Cybern.* **1971**, *1*, 71–96. [CrossRef]
7. Kauffman, S.A. Autocatalytic sets of proteins. *J. Theor. Biol.* **1986**, *119*, 1–24. [CrossRef]
8. Farmer, J.D.; Kauffman, S.A.; Packard, N.H. Autocatalytic replication of polymers. *Phys. D* **1986**, *22*, 50–67. [CrossRef]
9. Bagley, R.J.; Farmer, J.D. Spontaneous emergence of a metabolism. In *Artificial Life II*; Langton, C.G., Taylor, C., Farmer, J.D., Rasmussen, S., Eds.; Addison-Wesley: Boston, MA, USA, 1991; pp. 93–140.
10. Bagley, R.J.; Farmer, J.D.; Fontana, W. Evolution of a metabolism. In *Artificial Life II*; Langton, C.G., Taylor, C., Farmer, J.D., Rasmussen, S., Eds.; Addison-Wesley: Boston, MA, USA, 1991; pp. 141–158.

11.　Sievers, D.; von Kiedrowski, G. Self-replication of complementary nucleotide-based oligomers. *Nature* **1994**, *369*, 221–224. [CrossRef] [PubMed]

12.　Ashkenasy, G.; Jegasia, R.; Yadav, M.; Ghadiri, M.R. Design of a directed molecular network. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10872–10877. [CrossRef] [PubMed]

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

14.　Miras, H.N.; Mathis, C.; Xuan, W.; Long, D.L.; Pow, R.; Cronin, L. Spontaneous formation of autocatalytic sets with self-replicating inorganic metal oxide clusters. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 10699–10705. [CrossRef] [PubMed]

15.　Sousa, F.L.; Hordijk, W.; Steel, M.; Martin, W.F. Autocatalytic sets in *E. coli* metabolism. *J. Syst. Chem.* **2015**, *6*, 4. [CrossRef] [PubMed]

16.　Xavier, J.C.; Hordijk, W.; Kauffman, S.A.; Steel, M.; Martin, W.F. Autocatalytic chemical networks at the origin of metabolism. *Proc. R. Soc.* **2020**, *287*, 20192377. [CrossRef] [PubMed]

17.　Xavier, J.C.; Kauffman, S.A. Small-molecule autocatalytic networks are universal metabolic fossils. *Philos. Trans. R. Soc.* **2022**, *380*, 20210244. [CrossRef] [PubMed]

18.　Hordijk, W.; Steel, M. Chasing the tail: The emergence of autocatalytic networks. *BioSystems* **2017**, *152*, 1–10. [CrossRef] [PubMed]

19.　Serra, R.; Villani, M. *Modelling Protocells*; Springer: Berlin, Germany, 2017.

20.　Steel, M.; Hordijk, W.; Kauffman, S.A. Dynamics of a birth-death process based on combinatorial innovation. *J. Theor. Biol.* **2020**, *491*, 110187. [CrossRef] [PubMed]

21.　Hordijk, W.; Kauffman, S.A. Emergence of autocatalytic sets in a simple model of technological evolution. *arXiv* **2022**, arXiv:2204.01059

22.　Koppl, R.; Devereaux, A.; Herriot, J.; Kauffman, S. A simple combinatorial model of world economic history. *arXiv* **2018**, arXiv:1811.04502

23.　Hordijk, W.; Steel, M. Autocatalytic sets in polymer networks with variable catalysis distributions. *J. Math. Chem.* **2016**, *54*, 1997–2021. [CrossRef]