

The Influence of Modularity, Seeding, and Product Inhibition on Peptide Autocatalytic Network Dynamics

Wim Hordijk* Shira Shichor† Gonen Ashkenasy†

Abstract

Chemical networks often exhibit emergent, systems-level properties that cannot be simply derived from the linear sum of the individual components. The design and analysis of increasingly complex chemical networks thus constitute a major area of research in Systems Chemistry. In particular, much research is focused on the emergence of functional properties in prebiotic chemical networks relevant to the origin and early evolution of life. Here, we apply a formal framework known as RAF theory to study the dynamics of a complex network of mutually catalytic peptides. We investigate in detail the influence of network modularity, initial template seeding, and product inhibition on the network dynamics. We show that these results can be useful for designing new experiments, and further argue how they are relevant to origin of life studies.

1 Introduction

The origin of life inevitably involved the selection of functional molecules and self-organized molecular networks, made possible through the reliable replication of either individual molecules or entire sets of molecules. Minimal molecular self-replication has been demonstrated using synthetic nucleic acid and peptide entities [1, 2, 3, 4, 5, 6], and subsequently the formation of small replication networks has been realized experimentally [7, 8, 9].

One of the few successful examples of the latter involves a set of nine peptide molecules that mutually catalyze each other’s formation from shorter peptide fragments, such that the set as a whole is self-reproducing [10]. However, such experiments are often difficult to perform in the laboratory. Furthermore, it is difficult to design additional experiments necessary to reveal the intricate network dynamics and possible sets of inter-molecular sub-networks.

Networks of molecules often behave in complex and unpredictable ways, making it difficult to choose the right experimental ingredients and conditions.

*Institute for Advanced Study, University of Amsterdam, The Netherlands

†Ben-Gurion University of the Negev, Be’er Sheva, Israel

In systems biology, the use of computational modeling to investigate complex biological processes has been a common practice for many years [11]. Here, we apply a computational modeling approach to investigate complex chemical networks.

In particular, we apply the formal framework of autocatalytic sets to the above-mentioned self-reproducing peptide network. An *autocatalytic set* is a network of chemical reactions in which the molecules mutually catalyze each other’s formation from a basic food source. Thus, the self-reproducing peptide network serves as an experimental example of such an autocatalytic set. The general concept of autocatalytic sets has been formalized and studied in detail as RAF theory [12].

We use RAF theory to study the peptide autocatalytic set, investigating in particular the influence of network modularity, initial seeding, and product inhibition on the network dynamics and product distribution. We argue that the results of such a theoretical and computational investigation can be useful for designing new experiments. Indeed, recent experiments with the peptide network already verify some of the results obtained from the formal analysis presented here. Finally, we discuss how these results are relevant to origin of life studies.

2 Background

2.1 Autocatalytic sets

The concept of autocatalytic sets was originally introduced by Kauffman [13, 14, 15]. An *autocatalytic set* is defined as a set \mathcal{R} of reactions (and the molecule types involved in those reactions) that is:

1. *Reflexively autocatalytic* (RA): each reaction in \mathcal{R} is catalyzed by at least one of the molecule types involved in \mathcal{R} ; and
2. *F-generated* (F): each molecule type involved in \mathcal{R} can be created from the food set F through some sequence of reactions from \mathcal{R} .

The food set F is a (sub)set of molecule types that can be assumed to be directly available from the environment, i.e., they do not necessarily need to be produced through reactions in \mathcal{R} itself. A simple example, represented as a reaction graph, of an autocatalytic set is shown in Figure 1.

This notion of autocatalytic sets, or *RAF sets*, has been defined mathematically more rigorously, and an efficient (polynomial-time) algorithm for detecting RAF sets in general chemical reaction networks has also been developed [16, 17, 18, 12]. This RAF algorithm has been applied extensively to simple polymer-based computational models of chemical reaction networks, showing that autocatalytic sets are highly likely to exist at chemically realistic levels of catalysis, and under a wide variety of model assumptions [16, 17, 19, 20, 21, 22]. Moreover, several experimental self-replicating networks constructed in the lab have confirmed the chemical feasibility of RAF sets [23, 24, 10, 25].

Furthermore, computational studies have shown that RAF sets often consist of a hierarchy of smaller and smaller RAF subsets, i.e., smaller subsets of reactions that themselves are autocatalytic [26, 18]. Of particular interest, especially from a dynamical point of view, are the so-called *closed* RAF subsets. A closed RAF \mathcal{R}' is a subset of a RAF \mathcal{R} that contains all reactions $r \in \mathcal{R}$ for which all reactants and at least one catalyst are either generated by some reaction from \mathcal{R}' or are part of the food set F [21]. These closed RAFs represent the dynamically stable subsets of a RAF, i.e., they do not produce any new molecule types outside of those involved in \mathcal{R}' itself without the need for (rare) spontaneous, i.e. uncatalyzed, reactions.

Finally, the formal RAF framework has been applied successfully to analyze real chemical and biological reaction networks [27, 28].

2.2 The peptide autocatalytic set

More than a decade ago, Ashkenasy, Ghadiri, and colleagues constructed an autocatalytic set using nine helical peptides [10, 7]. This experimental autocatalytic set is represented in Figure 2, where the reaction products/templates (\mathbf{T}_i , $i = 1, \dots, 9$) are represented by the nodes in the graph, and the edges indicate which templates catalyze the formation of which products. The labels on the edges represent the relative template-product complex stability and can be associated with the efficiency of the respective catalytic processes (explained in more detail below).

Each reaction product \mathbf{T}_i is directly produced from the food set (not shown in the figure), which consists of the peptide fragments \mathbf{N} (a nucleophile) and \mathbf{E}_i (electrophiles). In other words, the reactions underlying the network are all of the form $\mathbf{N} + \mathbf{E}_i \rightarrow \mathbf{T}_i$. Thus, any subset of products \mathbf{T}_i is automatically food-generated. Furthermore, since each of the nine peptides has its formation catalyzed by at least one other peptide (i.e., there is at least one incoming edge for each node), this set is also reflexively autocatalytic, and thus forms a RAF set for the given food set. Since there are 25 edges in the graph, the full reaction network consists of 25 catalyzed reactions of the form $\mathbf{N} + \mathbf{E}_i + \mathbf{T}_j \rightarrow \mathbf{T}_i + \mathbf{T}_j$.

3 Methods

To perform stochastic dynamic simulations of this peptide autocatalytic set, the standard Gillespie algorithm is used [29, 30]. Rate constants for the 25 catalyzed reactions were calculated by extrapolation from the values in Table 1 (reproduced from [10]).

This table lists the theoretical (calculated) scores $s = -\Delta\Delta G$ of the template-product complexes, and experimental (measured) relative rate constants $r = K_{\text{rel}}$ for each of the nine autocatalytic reactions in isolation, as reported in [10]. Briefly, $-\Delta\Delta G$ compares the stability of all template-product complexes with that of a reference peptide ($-\Delta\Delta G = 0$) in which the mutated residues

	\mathbf{T}_1	\mathbf{T}_2	\mathbf{T}_3	\mathbf{T}_4	\mathbf{T}_5	\mathbf{T}_6	\mathbf{T}_7	\mathbf{T}_8	\mathbf{T}_9
s	3.0	6.9	4.8	2.4	8.4	4.5	4.8	6.9	4.2
r	1.0	22	7.1	<1	69	2.5	2.5	22	1.6

Table 1: Relative autocatalytic efficiency of the nine peptides in isolated reactions. Reproduced from [10].

had the neutral Alanine side chains. K_{rel} reflects a variation of the Michaelis-Menten evaluation of the autocatalytic reaction. K_{rel} was calibrated against the $K_{\text{cat}}/K_{\text{uncat}}$ of peptide \mathbf{T}_1 ; this ratio has been arbitrarily assigned with $K_{\text{rel}} = 1$. See [10] for details.

Using these s and r values, a quadratic relationship was fitted between them (see Figure 3):

$$r = 52.9 - 26.5s + 3.3s^2 \quad (1)$$

Next, using this fitted quadratic equation, relative rate constants r for all 25 catalyzed reactions were calculated using the corresponding scoring values s indicated on the edges in Figure 2, resulting in r values ranging from 5.5 for the slowest to 66.0 for the fastest reaction. In addition, each reaction $\mathbf{N} + \mathbf{E}_i \rightarrow \mathbf{T}_i$ is also allowed to happen spontaneously (i.e., uncatalyzed) with a lower rate constant $r_{sp} = 1.0$. So, there is a total of 25 catalyzed + 9 spontaneous = 34 reactions in the simulations.

Each simulation run is initialized with 50,000 molecules for each of the nine \mathbf{E}_i and 300,000 molecules of \mathbf{N} . This ratio between the \mathbf{E}_i and \mathbf{N} initial concentrations results in competition between the different \mathbf{E}_i to react with an under-stoichiometric amount of \mathbf{N} , as also applied in the original experiments [10]. Each simulation is then run for a given amount of time T until most of the food molecules (peptide fragments \mathbf{E}_i and \mathbf{N}) have been converted into full-length peptides (products \mathbf{T}_i).

Since the relative rate constants r in the simulation have no explicit units, we need to choose appropriate values for the volume V and the simulation time T . We have chosen $V = 10^{11}$ (again in arbitrary units), in which case all food molecules are converted into products within $T = 10$ (arbitrary) units of time.

Finally, to determine the closed RAF subsets within a given RAF set \mathcal{R} , the chemical organizations [31] of \mathcal{R} are computed, which include all its closed RAF subsets [32].

4 Results

4.1 Modularity

The 9-peptide RAF set \mathcal{R} in Figure 2 contains several smaller subsets that form RAF sets by themselves, such as for example the subset $\{\mathbf{T}_3, \mathbf{T}_5, \mathbf{T}_7\}$, since these molecules all catalyze each other’s formation. However, only two of these RAF subsets are relevant from a dynamical point of view, forming closed RAFs.

First note that no RAF subset can be realized (dynamically) without at least one spontaneous (i.e., uncatalyzed) reaction. When starting the simulation with the food set only, none of the catalysts (the products \mathbf{T}_i) are present yet, so at least one of them will have to be produced by one of the reactions $\mathbf{N} + \mathbf{E}_i \rightarrow \mathbf{T}_i$ happening spontaneously. This is of course always possible, but will happen at a lower rate than a catalyzed reaction. But once one catalyst is produced spontaneously, it can start catalyzing the production of other products (i.e., catalysts), and different RAF subsets can come into existence depending on which catalyst was produced first.

In particular, if one of the catalysts \mathbf{T}_1 , \mathbf{T}_3 , \mathbf{T}_4 , \mathbf{T}_5 , or \mathbf{T}_7 is produced spontaneously, then the RAF subset $\mathcal{R}_1 = \{\mathbf{T}_1, \mathbf{T}_3, \mathbf{T}_4, \mathbf{T}_5, \mathbf{T}_7, \mathbf{T}_9\}$ can come into existence, due to the high level of mutual catalysis within this subset. However, none of the peptides in the subset \mathcal{R}_1 catalyzes any of the peptides in the subset $\mathcal{R}_2 = \{\mathbf{T}_2, \mathbf{T}_6, \mathbf{T}_8\}$. So, the only way to get the RAF subset \mathcal{R}_2 to come into existence is to have one of its members produced spontaneously. When this happens, though, not only \mathcal{R}_2 itself, but also \mathcal{R}_1 can come into existence, since the members of \mathcal{R}_2 do catalyze several members of \mathcal{R}_1 .

Consequently, given that $\mathcal{R}_1 \cup \mathcal{R}_2 = \mathcal{R}$, the only two *closed* RAF subsets are \mathcal{R}_1 and the full RAF set \mathcal{R} . This is easily verified by computing the chemical organizations within \mathcal{R} , confirming this modularity (i.e., existence of more than one closed RAF subset).

Finally, note that if \mathbf{T}_9 is produced spontaneously, still no RAF subset can come into existence, since \mathbf{T}_9 does not catalyze the production of any other peptide, but it is catalyzed by one member of both \mathcal{R}_1 and \mathcal{R}_2 .

To show the influence of this modularity on the possible dynamics of the system, Gillespie simulations were performed on the 9-peptide reaction network. Figure 4 shows an example where \mathbf{T}_2 and \mathbf{T}_6 were produced through spontaneous reactions, giving rise to the closed RAF \mathcal{R} , i.e., the full 9-peptide RAF set.

Note that in this case \mathbf{T}_5 , \mathbf{T}_7 , \mathbf{T}_1 , and \mathbf{T}_3 are produced in the highest quantities, \mathbf{T}_4 and \mathbf{T}_9 in intermediate quantities, and \mathbf{T}_2 , \mathbf{T}_8 , and \mathbf{T}_6 in the lowest quantities. Most of the members of \mathcal{R}_1 are being produced in the highest quantities, not only due to their own high level of mutual catalysis, but also because some of them are being catalyzed by members of \mathcal{R}_2 . On the other hand, the members of \mathcal{R}_2 are being produced in the lowest quantities, as they only mutually catalyze each other but are not catalyzed by any members of \mathcal{R}_1 .

Figure 5 shows another example, but one where \mathbf{T}_3 and \mathbf{T}_4 were produced by spontaneous reactions, in which case only \mathcal{R}_1 is produced. Since none of the members of \mathcal{R}_2 were produced spontaneously, this subset never came into existence (recall that none of its members are catalyzed by \mathcal{R}_1).

As a third possibility, if one of the members of \mathcal{R}_2 is produced spontaneously but at a later time than one of the members of \mathcal{R}_1 , then most of the food molecules will be converted into members of \mathcal{R}_1 , in which case the concentrations of \mathcal{R}_2 remain very low. An example of this type of dynamics is shown in Figure 6 where \mathbf{T}_8 was created through a spontaneous reaction sometime after \mathbf{T}_1 had already been created.

In conclusion, the network modularity, i.e., the existence of two closed RAF subsets, has a clear influence on the possible dynamics of the system. Depending on which catalysts are produced through initial spontaneous reactions, and in which order, at least three different dynamical behaviors (attractors) are possible.

These dynamical simulations reflect the theoretical network analysis, in particular the existence of two separate subsets \mathcal{R}_1 and \mathcal{R}_2 . Looking at the final distribution of peptides (at the end of the simulation), one finds that the experimental results from [10] were partially reproduced. Using the HPLC data shown in Fig. 3 of the original paper [10], which is representative of several experimental repetitions, and comparing them with the final distribution of the simulation run shown in Figure 4, the “relative abundances” of the nine peptides are compared in Figure 7.

As this histogram shows, there are some differences between the two distributions. For example, whereas \mathbf{T}_8 has the highest relative abundance (about 15%) in the experiment, it has one of the lowest relative abundances (about 5%) in the simulation. On the other hand, in the simulation \mathbf{T}_3 and \mathbf{T}_5 end up at one of the highest relative abundances (15% or higher), whereas these have some of the lower relative abundances in the experiment (8-9%). Even though there are fluctuations in these abundances between repetitions of the same experiment, these differences between the experimental and simulation results are larger than what could be expected from mere experimental errors.

Moreover, the experimental results do not reflect the network modularity made explicit in the simulations. So, most likely there are some additional catalytic links in the experimental system that were not included in the theoretically calculated links in Figure 2. For example, Figure 8 shows the result of a simulation where \mathbf{T}_1 also catalyzes \mathbf{T}_2 and \mathbf{T}_8 (with a threshold score of 5.6), making it a “fully connected” network, i.e., there is no modularity (separate closed RAF subsets) anymore. In this case, \mathbf{T}_2 and \mathbf{T}_8 are produced in larger quantities than before (i.e., compared with Figure 4), bringing the simulation results much closer to the experimental result.

For the simulations described in the following subsections, the fully connected version of the network is used (with \mathbf{T}_1 also catalyzing \mathbf{T}_2 and \mathbf{T}_8), to make sure the simulation results reflect more accurately the observed experimental results. Furthermore, in the discussion section we describe how environmental conditions can be altered to induce modularity explicitly, also in the actual experimental systems.

4.2 Seeding

As already became clear in the previous section, the relative abundances with which the various peptides are produced also depend sensitively on which peptides are produced initially by spontaneous reactions. To make this dependence more explicit, we repeated the simulations but seeding them with 50 molecules of one particular \mathbf{T}_i (once for each of the nine peptides), rather than allowing spontaneous reactions to happen. In these simulations, we use the fully con-

nected network, i.e., with \mathbf{T}_1 catalyzing the formation of \mathbf{T}_2 and \mathbf{T}_8 (with a threshold score of 5.6), for a more fair comparison with experimental results.

Figure 9 shows the relative abundances for all nine peptides for each possible seed. As this plot shows, for some peptides it does not matter much what the seed was. For example, \mathbf{T}_1 , \mathbf{T}_5 , and \mathbf{T}_9 are relatively insensitive to the type of seed. However, especially \mathbf{T}_2 , \mathbf{T}_6 , and \mathbf{T}_8 are quite sensitive to the seed. This plot also confirms the earlier claim that if the system is seeded with \mathbf{T}_9 , none of the peptides are produced, given that \mathbf{T}_9 does not catalyze any reactions.

This dependency of production rates for each of the nine peptides on the seed was also demonstrated with the chemical experiments [10], and is clearly reflected in the simulations as well.

4.3 Product inhibition

In the simulations so far, it has been assumed that the catalyst and the newly formed product dissociate immediately after the ligation reaction has happened. However, in reality this dissociation happens only at a given (often slow) rate, which may result in *product inhibition*, i.e., the newly formed product is not immediately available to act as a catalyst itself. To investigate the effect of product inhibition on the network dynamics, the catalyzed $\mathbf{N} + \mathbf{E}_i \rightarrow \mathbf{T}_i$ reactions are considered to happen in two steps:

1. $\mathbf{N} + \mathbf{E}_i + \mathbf{T}_j \rightarrow \mathbf{T}_i \bullet \mathbf{T}_j$
2. $\mathbf{T}_i \bullet \mathbf{T}_j \rightarrow \mathbf{T}_i + \mathbf{T}_j$

The reactions in the first step have the same rate constants r as the original (catalyzed) ligation reactions, while the dissociation reactions in the second step now have rate constants that are inversely proportional to the corresponding r values.

In particular, the dissociation reactions have rate constants of the form $k \times 0.01 \times 10^a$, where k is determined as follows. Recall from the Methods that the rate constants r for the (catalyzed) ligation reactions range from 5.5 to 66.0. Dividing this range into ten roughly equally spaced bins from high (66.0) to low (5.5), if the rate constant r of a ligation reaction $\mathbf{N} + \mathbf{E}_i + \mathbf{T}_j \rightarrow \mathbf{T}_i \bullet \mathbf{T}_j$ falls in the k^{th} highest bin, then that value of k is used in calculating the rate constant for the corresponding dissociation reaction $\mathbf{T}_i \bullet \mathbf{T}_j \rightarrow \mathbf{T}_i + \mathbf{T}_j$. In other words, a more efficient catalyst (higher value of r) is assumed to have a lower dissociation rate.

Finally, the parameter a reflects the strength of the product inhibition. Here, $a = 0, -1, -2, -3, -4$ are used, where $a = 0$ results in weak product inhibition (high rate of dissociation) and $a = -4$ results in very strong product inhibition (low rate of dissociation). Each simulation run is seeded with 10 molecules of \mathbf{T}_1 (no spontaneous reactions are allowed), and the fully connected network is again used (including \mathbf{T}_1 catalyzing the formation of \mathbf{T}_2 and \mathbf{T}_8).

Figure 10 shows the result of a simulation with $a = 0$ (i.e., weak product inhibition). However, the effect is already significant. First, there is a longer

lag time (about 15 time units rather than just 1 or 2) before the peptides start becoming available (the “onset” of product formation). Second, the relative abundances are also different compared to the original simulation run without product inhibition (refer to Figure 8).

Figure 11 shows the effect of the strength of product inhibition on the (initial) growth rate of one of the products (\mathbf{T}_1). For decreasing values of a (i.e., the order of magnitude by which the dissociation rate is decreased), the growth rate of \mathbf{T}_1 decreases significantly. The curves in Figure 11 show the amount of product only from the onset of product formation (i.e., after the lag time). However, the actual lag time for this onset itself increases about one order of magnitude with each decrease of one order of magnitude in the parameter a .

Finally, the strength of product inhibition also has a significant effect on the relative abundances of the products. Figure 12 shows histograms of the relative product abundances at the end of each simulation run for different values of the parameter a . Only \mathbf{T}_2 and \mathbf{T}_8 do not seem to be influenced, but for the others there is either a clear decrease or increase with increasing strength of product inhibition (i.e., decreasing a).

In fact, the effect of product inhibition brings some of the relative abundances even closer to the experimental results, suggesting that product inhibition does indeed play a role in the experimental system.

5 Discussion

We have shown how the dynamics of a particular autocatalytic network consisting of peptides depends on at least three topological and mechanistic factors:

1. *Modularity* of the reaction network (i.e., existence of multiple closed RAFs and their cross-talk).
2. *Seeding* of the system (i.e., which catalyst is introduced or spontaneously produced first).
3. *Product inhibition* (i.e., dissociation rate between a catalyst and the newly formed product).

Network modularity has been shown to be crucial for the potential evolvability of autocatalytic sets [33, 20, 34]. Such modularity clearly exists within the theoretically calculated peptide network, both in terms of the network analysis (in the form of closed RAFs) and in the dynamical simulations (in the form of different possible relative abundances of the various peptides).

The 9-peptide RAF set studied here was originally taken from a larger set of 25 peptides for which catalytic connections (scores) were calculated theoretically [10]. This full 25-peptide set is reproduced in Figure 13, and also forms a RAF set. Moreover, this RAF set contains many RAF subsets, including 12 closed RAFs (determined by computing the chemical organizations within the full RAF set). These closed RAFs are represented by the colored circles and the various possible combinations of them.

Given that the full 25-peptide RAF set has 12 closed RAF subsets, compared to two for the 9-peptide network, an even wider range of possible dynamics can be expected from this larger network. Figure 14 shows the results of two simulation runs on the 25-peptide network, showing the number of molecules over time for one of the members of each closed RAF in Figure 13 (using corresponding colors). Initial conditions and reaction rate constants are similar to those of the simulations on the 9-peptide networks: 50,000 molecules of each of the 25 \mathbf{E}_i , 850,000 molecules of \mathbf{N} (about 2/3 of the total amount of \mathbf{E}_i), and converting the calculated scores s to rate constants r using the estimated quadratic relation. As the plots show, which closed RAFs come into existence first (or at all) depends on which of the catalysts are created (by chance) through initial spontaneous reactions.

We further argue that using the detailed insights gained from the theoretical network analysis and computational simulations, it should be possible to construct alternative experimental peptide networks and/or to find conditions that will explicitly alter the modularity. An experiment in this direction was already performed by Ashkenasy et al., where varying the environmental conditions through lowering the medium pH or increasing salt concentration showed significant network rewiring compared to the “native” reaction conditions (see Figure 15), giving rise to a new set of closed RAF subsets and selective amplification of one or more of the RAF products [35]. Lowering the pH results in the formation of a new dominant RAF subset $\{\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_7\}$, in which the three peptides support each other’s formation and are also formed through additional catalytic pathways. Increasing the salt concentration led to a unique RAF, in which the formation of a central node \mathbf{T}_4 is supported by multiple catalytic pathways.

If a given network does indeed possess modularity, then stochastic events such as spontaneous reactions can directly affect the overall dynamics. However, these alternative dynamics can be explicitly controlled by seeding the system in specific ways, rather than relying on spontaneous reactions. By manipulating the peptide networks discussed here (or their close analogs) it has thus been shown, both experimentally and theoretically, that it is possible to perform (simple) computations and network motifs [36, 37, 38, 39, 40].

It has been shown in actual chemical experiments, and also in theory, that product inhibition can influence the overall behavior of an autocatalytic system, in particular the (initial) product growth rates being parabolic instead of exponential [41, 2, 7]. Such an effect can potentially give rise to the coexistence of different molecular replicators, rather than one outcompeting all others [42]. Here we have shown explicitly by simulations how product inhibition affects the behavior of an *entire autocatalytic network*, not only in terms of the (initial) growth rates, but also in terms of the relative abundances of the various peptides.

Finally, we argue that since the current simulations were applied to analyze a realistic experimental system, composed of molecules of biological origin (the coiled coil peptide motifs), our results provide further support for the validity of the formal autocatalytic sets (RAF) framework for describing scenarios relevant

to the emergence of functional molecules in early chemical evolution, potentially leading to the (or an) origin of life. Furthermore, the RAF analysis revealed the flexibility in network topology and dynamics potentially available within peptide networks, re-iterating their likely involvement in pre-cellular molecular evolution [15, 43, 44, 45, 46], prior to or in co-existence with an RNA world.

6 Acknowledgments

WH thanks the Institute for Advanced Study, University of Amsterdam, for financial support in the form of a fellowship. BGU research is supported by the Israeli Science Foundation (ISF 707/16). We also thank COST Action CM1304 for funding WH's research visit at BGU, from which this paper resulted.

References

- [1] O. Taran and G. von Kiedrowski. Replicators: Components for systems chemistry. In P. L. Luisi and C. Chiarabelli, editors, *Chemical Synthetic Biology*, pages 289–319, 2011.
- [2] N. Paul and G. F. Joyce. Minimal self-replicating systems. *Current Opinion in Chemical Biology*, 8:634–639, 2004.
- [3] A. Vidonne and D. Philp. Making molecules make themselves – the chemistry of artificial replicators. *European Journal of Organic Chemistry*, 2009(5):593–610, 2009.
- [4] A. J. Bissette and S. P. Fletcher. Mechanisms of autocatalysis. *Angewandte Chemie International Edition*, 52(49):12800–12826, 2013.
- [5] G. Danger, R. Plasson, and R. Pascal. Pathways for the formation and evolution of peptides in prebiotic environments. *Chemical Society Reviews*, 41:5416–5429, 2012.
- [6] M. Stich, J. M. Ribo, D. G. Blackmond, and D. Hochberg. Necessary conditions for the emergence of homochirality via autocatalytic self-replication. *Journal of Chemical Physics*, 145(7):074111, 2016.
- [7] Z. Dadon, N. Wagner, and G. Ashkenasy. The road to non-enzymatic molecular networks. *Angewandte Chemie International Edition*, 47:6128–6136, 2008.
- [8] T. Kosikova and D. Philp. Exploring the emergence of complexity using synthetic replicators. *Chemical Society Reviews*, 46:7274–7305, 2017.
- [9] J. Li, P. Nowak, and S. Otto. Dynamic combinatorial libraries: From exploring molecular recognition to systems chemistry. *Journal of the American Chemical Society*, 135(25):9222–9239, 2013.

- [10] G. Ashkenasy, R. Jegasia, M. Yadav, and M. R. Ghadiri. Design of a directed molecular network. *PNAS*, 101(30):10872–10877, 2004.
- [11] E. Bartocci and P. Lio. Computational modeling, formal analysis, and tools for systems biology. *PLOS Computational Biology*, 12(1):e1004591, 2016.
- [12] W. Hordijk and M. Steel. Chasing the tail: The emergence of autocatalytic networks. *BioSystems*, 152:1–10, 2017.
- [13] S. A. Kauffman. Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. *Journal of Cybernetics*, 1(1):71–96, 1971.
- [14] S. A. Kauffman. Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119:1–24, 1986.
- [15] S. A. Kauffman. *Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, 1993.
- [16] W. Hordijk and M. Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227(4):451–461, 2004.
- [17] 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.
- [18] W. Hordijk, J. I. Smith, and M. Steel. Algorithms for detecting and analysing autocatalytic sets. *Algorithms for Molecular Biology*, 10:15, 2015.
- [19] 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.
- [20] 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.
- [21] J. Smith, M. Steel, and W. Hordijk. Autocatalytic sets in a partitioned biochemical network. *Journal of Systems Chemistry*, 5:2, 2014.
- [22] W. Hordijk and M. Steel. Autocatalytic sets in polymer networks with variable catalysis distributions. *Journal of Mathematical Chemistry*, 54(10):1997–2021, 2016.
- [23] D. Sievers and G. von Kiedrowski. Self-replication of complementary nucleotide-based oligomers. *Nature*, 369:221–224, 1994.
- [24] D-E. Kim and G. F. Joyce. Cross-catalytic replication of an RNA ligase ribozyme. *Chemistry & Biology*, 11:1505–1512, 2004.

- [25] 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.
- [26] W. Hordijk, M. Steel, and S. Kauffman. The structure of autocatalytic sets: Evolvability, enablement, and emergence. *Acta Biotheoretica*, 60(4):379–392, 2012.
- [27] W. Hordijk and M. Steel. A formal model of autocatalytic sets emerging in an RNA replicator system. *Journal of Systems Chemistry*, 4:3, 2013.
- [28] F. L. Sousa, W. Hordijk, M. Steel, and W. F. Martin. Autocatalytic sets in *E. coli* metabolism. *Journal of Systems Chemistry*, 6:4, 2015.
- [29] 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.
- [30] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
- [31] P. Dittrich and P. Speroni di Fenizio. Chemical organization theory. *Bulletin of Mathematical Biology*, 69(4):1199–1231, 2007.
- [32] W. Hordijk, M. Steel, and P. Dittrich. Autocatalytic sets and chemical organizations: Modeling self-sustaining reaction networks at the origin of life. *New Journal of Physics*, 20:015011, 2018.
- [33] V. Vasas, C. Fernando, M. Santos, S. Kauffman, and E. Sathmáry. Evolution before genes. *Biology Direct*, 7:1, 2012.
- [34] W. Hordijk, J. Naylor, N. Krasnogor, and H. Fellermann. Dynamics of autocatalytic sets in populations of compartments. *Scientific Reports*, 2018. (under review).
- [35] Z. Dadon, N. Wagner, R. Cohen-Luria, and G. Ashkenasy. Reaction networks. In P. Gale and J. Steed, editors, *Supramolecular Chemistry: From Molecules to Nanomaterials*, volume 4, pages 1527–1542, 2012.
- [36] G. Ashkenasy and M. R. Ghadiri. Boolean logic functions of a synthetic peptide network. *Journal of the American Chemical Society*, 126:11140–11141, 2004.
- [37] N. Wagner and G. Ashkenasy. Systems chemistry: Logic gates, arithmetic units, and network motifs in small networks. *Chemistry - A European Journal*, 15:1765–1775, 2009.
- [38] N. Wagner and G. Ashkenasy. Symmetry and order in systems chemistry. *Journal of Chemical Physics*, 130:164907/1–164907/6, 2009.

- [39] Z. Dadon, M. Samiappan, E. Yishay, and G. Ashkenasy. Light-induced peptide replication controls logic operations in small networks. *Chemistry - A European Journal*, 16:12096–12099, 2010.
- [40] M. Samiappan, Z. Dadon, and G. Ashkenasy. Replication NAND gate with light as input and output. *Chemical Communications*, 47:710–712, 2011.
- [41] G. von Kiedrowski. Minimal replicator theory I: parabolic versus exponential growth. In D. H. Berlin, editor, *Bioorganic Chemistry*, volume 3, pages 115–146. Springer-Verlag, 1993.
- [42] E. Szathmáry and I. Gladkih. Sub-exponential growth and coexistence of non-enzymatically replicating templates. *Journal of Theoretical Biology*, 138:55–58, 1989.
- [43] L. E. Orgel. The origin of life—a review of facts and speculations. *Trends in Biochemical Sciences*, 23(12):491–495, 1998.
- [44] A. Brack. From interstellar amino acids to prebiotic catalytic peptides: A review. *Chemistry & Biodiversity*, 4(4):665–679, 2007.
- [45] K. Ruiz-Mirazo, C. Briones, and A. de la Escosura. Prebiotic systems chemistry: New perspectives for the origins of life. *Chemical Reviews*, 114(1):285–366, 2014.
- [46] P. L. Luisi, C. Chiarabelli, and P. Stano. From the never born proteins to the minimal living cell: Two projects in synthetic biology. *Origins of Life and Evolution of Biospheres*, 36:605–616, 2006.
- [47] M. Steel, W. Hordijk, and J. Smith. Minimal autocatalytic networks. *Journal of Theoretical Biology*, 332:96–107, 2013.

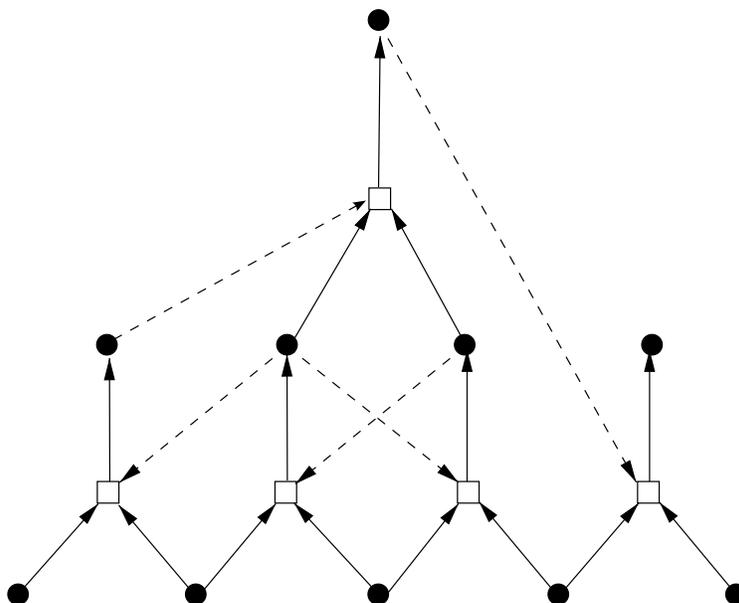


Figure 1: A simple example of an autocatalytic set \mathcal{R} consisting of five reactions. Dots represent molecule types, while boxes represent reactions, with solid arrows indicating reactants going into and products coming out of a reaction. Dashed arrows indicate catalysis. The food set F consists of the five molecule types at the bottom. Adapted from [47].

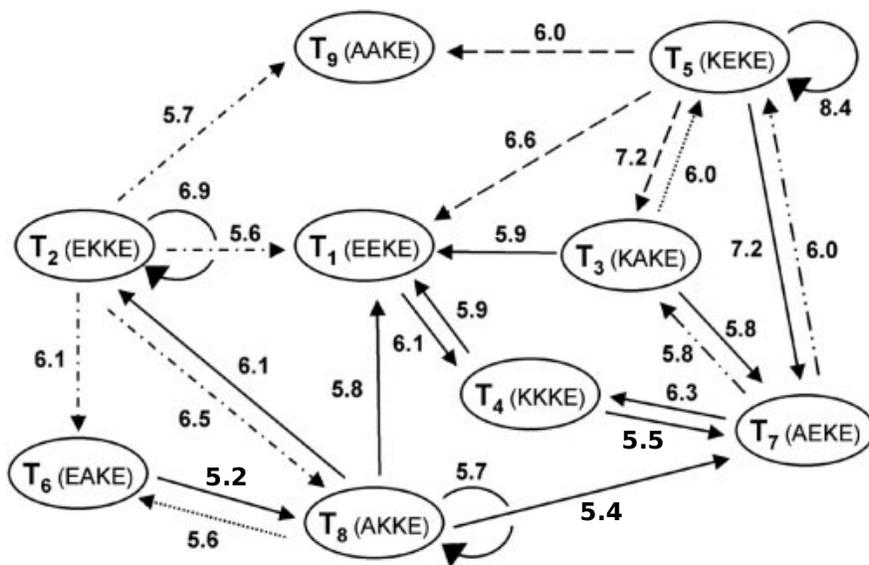


Figure 2: The experimentally characterized peptide-based autocatalytic set. Arrows indicate which peptides catalyze the formation of which others, with labels indicating the corresponding score. The different line styles (solid, dashed, or dotted) reflect different ways in which the scores were experimentally measured. Adapted from [10], with the three missing scores (for values < 5.6) added.

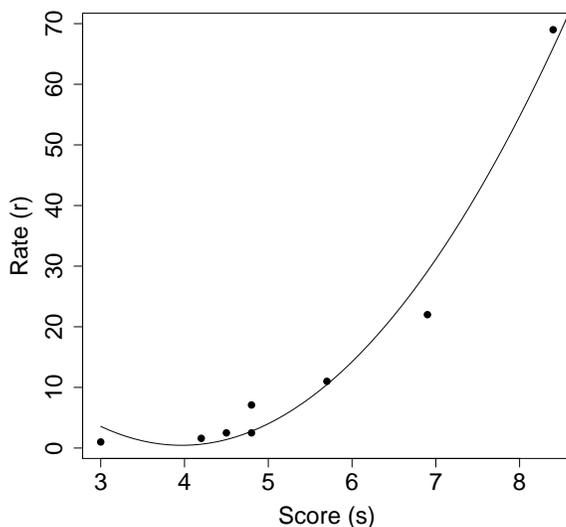


Figure 3: The quadratic fit (solid line) between the experimentally observed autocatalytic rate constants r and scoring values s from Table 1 (dots). The regression has an R^2 value of 0.976 (i.e., it is a very good fit), with all estimated parameter values being highly significant (i.e., a significance level of 0.01 or better).

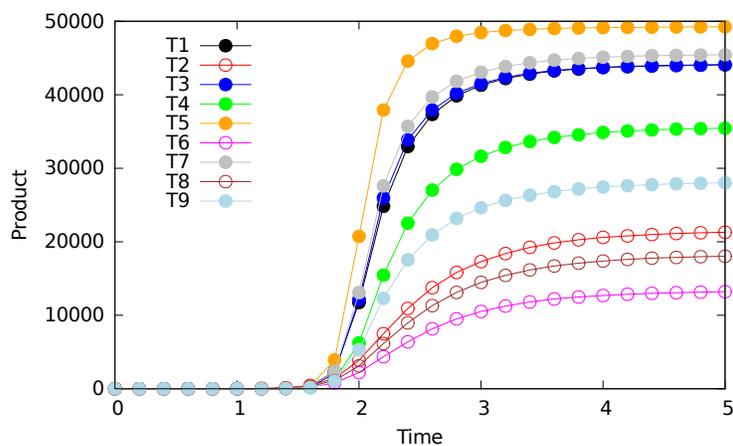


Figure 4: A simulation run with two spontaneous reactions first producing \mathbf{T}_2 and then \mathbf{T}_6 . Note that the subset \mathcal{R}_1 is indicated with solid dots and the subset \mathcal{R}_2 with open dots.

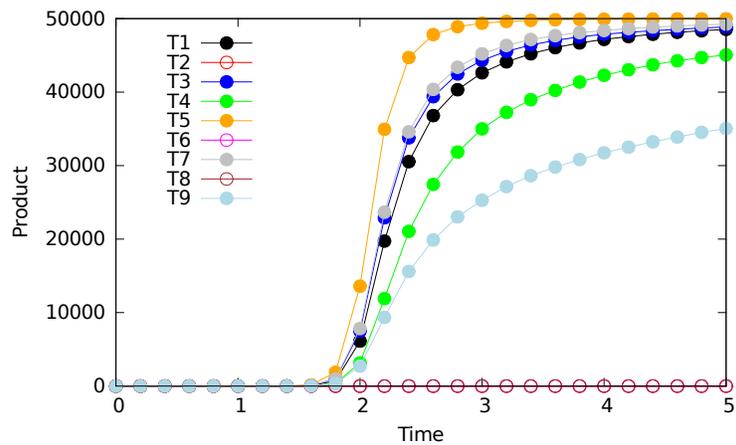


Figure 5: A simulation run with two spontaneous reactions first producing \mathbf{T}_3 and then \mathbf{T}_4 . Solid and open dots again distinguish \mathcal{R}_1 and \mathcal{R}_2 .

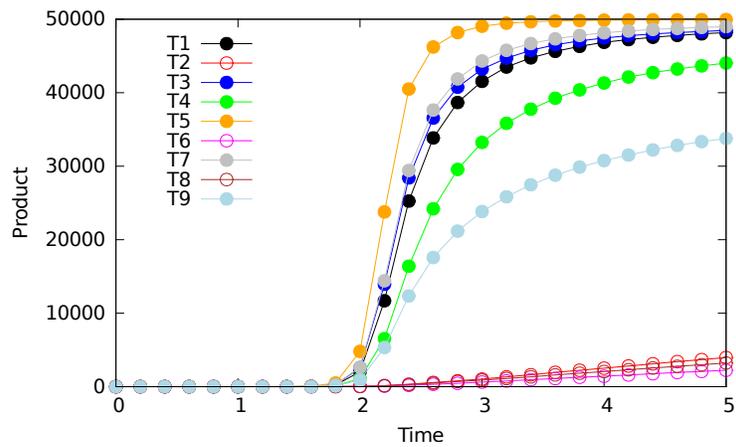


Figure 6: A simulation run with two spontaneous reactions producing \mathbf{T}_1 first and then \mathbf{T}_8 later on.

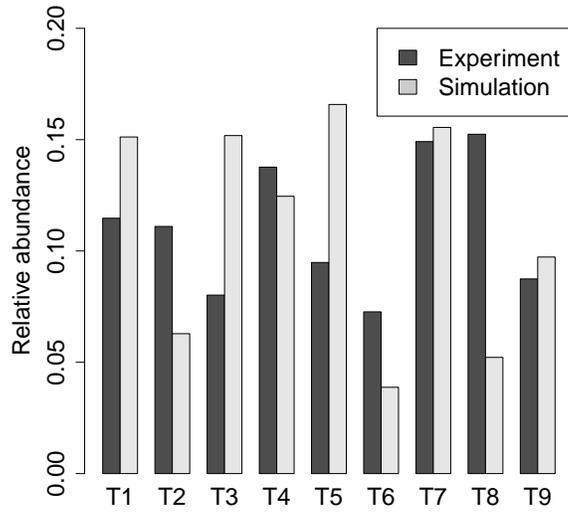


Figure 7: The relative abundance (in percentage of total concentration) of the nine peptides \mathbf{T}_i from an experiment (dark bars) and the simulation presented in Figure 4 (light bars).

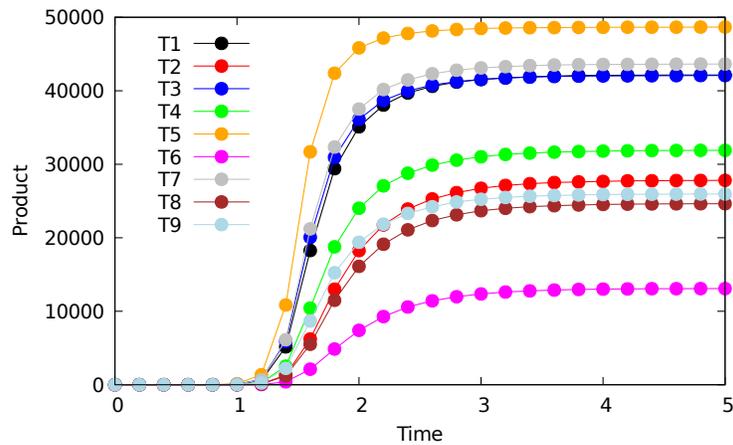


Figure 8: The fully connected network with two spontaneous reactions first producing \mathbf{T}_7 and then \mathbf{T}_5 . Since there is no modularity anymore, all products are indicated with solid dots.

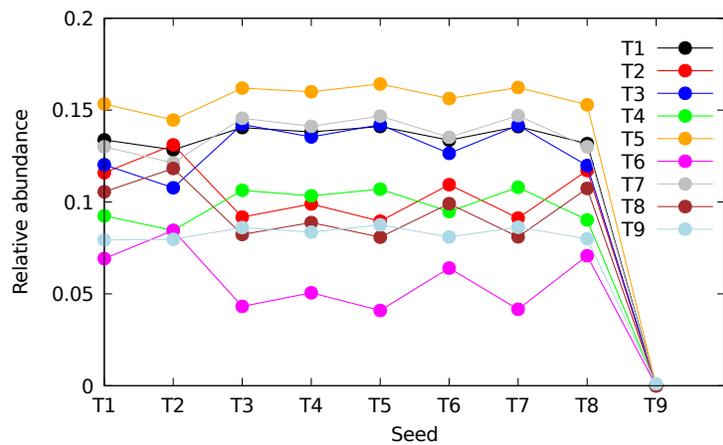


Figure 9: The relative abundances of all products at the end of a simulation, depending on which catalyst was introduced as the seed (50 molecules).

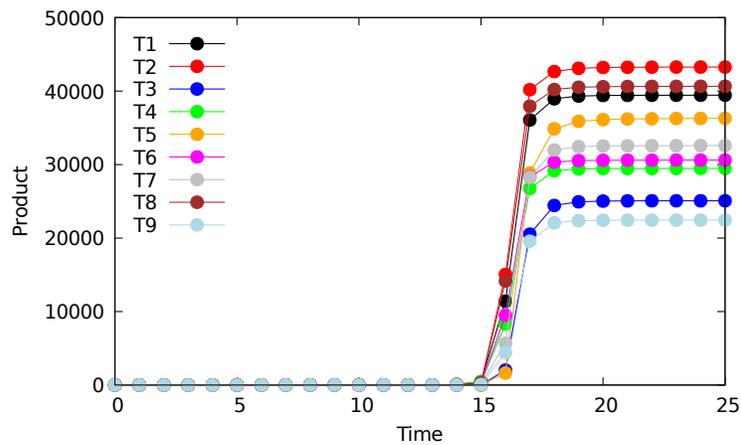


Figure 10: A simulation run with weak product inhibition, and initial seeding with 10 molecules of T_1 .

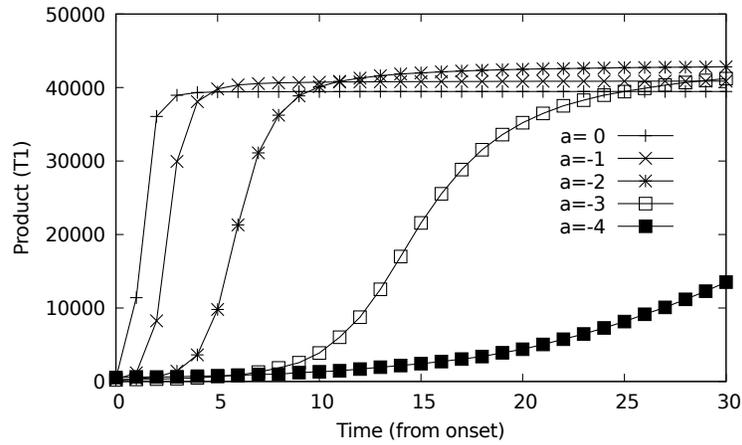


Figure 11: The (initial) growth of T_1 depending on the strength of product inhibition.

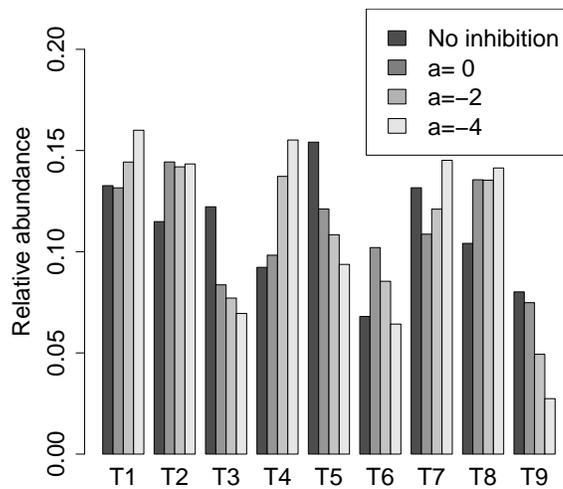


Figure 12: The relative abundances (in percentage of total concentration) of the nine peptides T_i for various strengths of product inhibition.

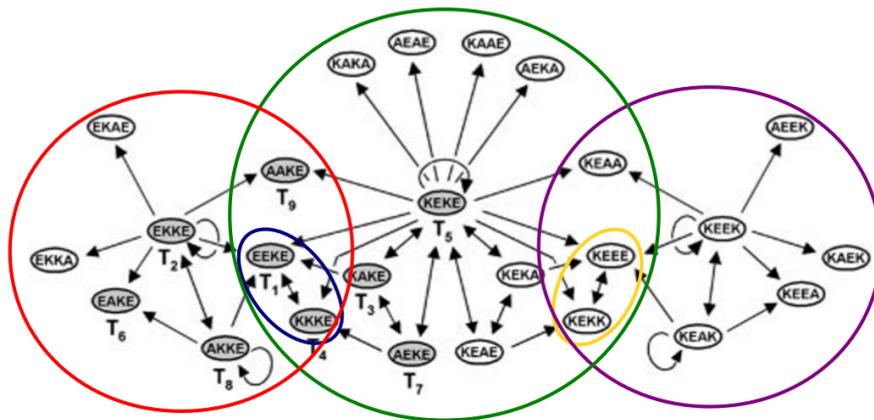


Figure 13: The full 25-peptide autocatalytic set and its closed RAF subsets (indicated by colored circles). Adapted from [10].

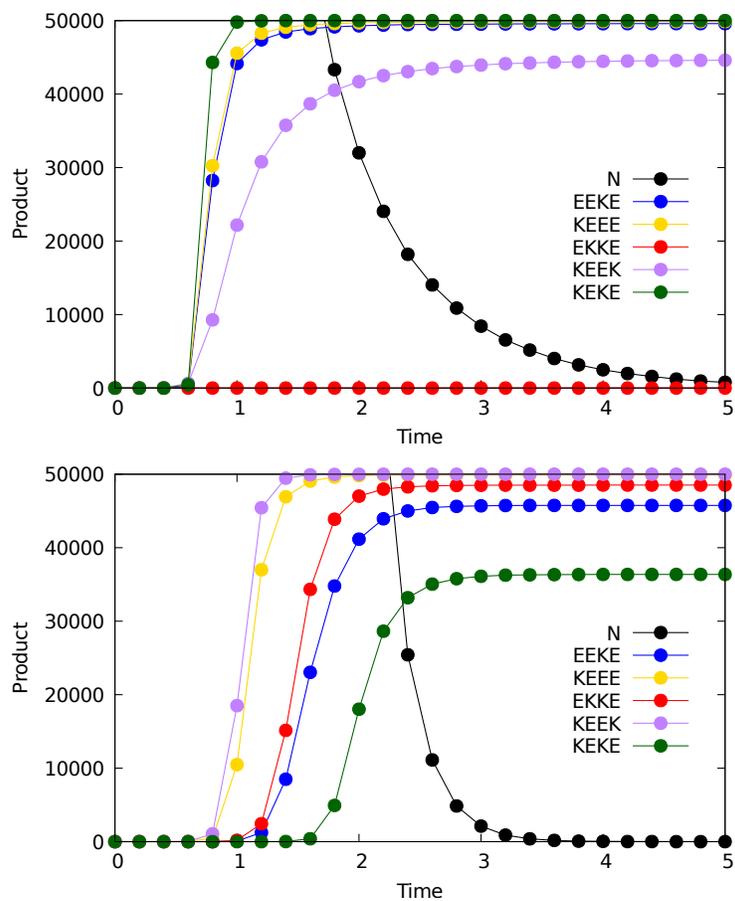


Figure 14: Two different simulation results on the same 25-peptide reaction network. The differences are purely due to stochastic events, i.e., different spontaneous (uncatalyzed) reactions happening initially. The black line shows the number of “food” (N) molecules left in the system over time.

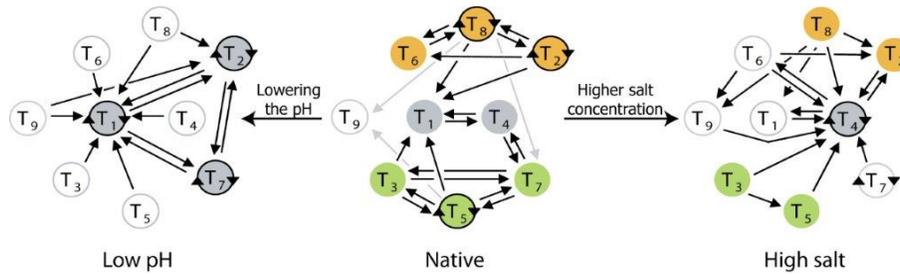


Figure 15: A 9-mer network connectivity in different environmental conditions. The different colors emphasize the sub-networks that were formed under these distinct settings. The network topology under 'Native' conditions (middle) is identical to that shown in Figure 2, but highlighting the RAF subsets slightly differently to allow better comparison with the networks under variable conditions (left and right). Adapted from [35].