Structure and selection in an autocatalytic binary polymer model

Shinpei Tanaka\(^1\), Harold Fellermann\(^2,3\) and Steen Rasmussen\(^2,4\)

\(^1\) Graduate School of Integrated Arts and Sciences, Hiroshima University - I-7-1 Kagamiyama, Higashi-Hiroshima 739-8521, Japan
\(^2\) Center for Fundamental Living Technology (FLinT) Department of Physics, Chemistry and Pharmacy, University of Southern Denmark - Campusvej 55 5230 Odense M, Denmark
\(^3\) School of Computing Science, Newcastle University - Claremont Tower, Newcastle upon Tyne, UK
\(^4\) Santa Fe Institute - 1399 Hyde Park Rd, Santa Fe, NM 87501, USA

received 17 March 2014; accepted in final form 20 June 2014
published online 14 July 2014

PACS 87.23.Kg – Dynamics of evolution
PACS 05.65.+b – Self-organized systems
PACS 87.23.Cc – Population dynamics and ecological pattern formation

Abstract – An autocatalytic binary polymer system is studied as an abstract model for a chemical reaction network capable to evolve. Due to autocatalysis, long polymers appear spontaneously and their concentration is shown to be maintained at the same level as that of monomers. When the reaction starts from a pool of monomers, highly ordered populations with particular sequence patterns are dynamically selected out of a vast number of possible states. The interplay between the selected microscopic sequence patterns and the macroscopic cooperative structures is examined both analytically and in simulation. Stability, fluctuations, and dynamic selection mechanisms are investigated for the involved self-organizing processes.

Introduction. – Prebiotic emergence of inheritable information hinges on the ability of nucleic acid sequences to act as templates for their own replication from a pool of monomers and shorter substrands [1]. If this copying process occurs with sufficient fidelity [1] and at a rate faster than degradation [2], long sequence information can be maintained over extended periods of time. In such a scenario, replicating sequences compete for common resources, and this competition could drive a selection process which, in turn, directs the random exploration process of copy error mutations. While the emergence of first prebiotic replicators is subject to ongoing debate [3–5], the resulting evolutionary search is often assumed to eventually select replicator species with some advantageous prebiotic properties [6]—thus opening up for Darwinian evolution.

Here, we analyze a simple mathematical replicator model and demonstrate that the competition process of template directed replicators does not necessarily lead to a free exploration of the search space, but rather the detailed reaction mechanism can indeed induce a strong bias on the selected sequence information. This intrinsic selection is due to symmetries in the replication process, and the selected species may or may not exhibit advantageous prebiotic properties such as secondary structure formation or catalytic activity.

Non-enzymatic DNA and RNA replication has been studied extensively in vitro [7,8], where it has been observed that replication from monomers performs with negligible yield [9], whereas replication of a template strand from shorter oligomer substrands can lead to significant amplification of the template species [7,10,11].

The emergence of autocatalytic networks in model chemistries has also been a prominent research subject in complex systems, artificial life and origins of life studies. Earlier work is mainly concerned with the emergence and dynamics of autocatalytic sets from random cross-catalytic molecules [12–18], or focused on the detailed dynamics of template directed replication systems [7,19,20].

In this study we simplify the reaction network of template directed replication from substrands as much as possible, in order to determine how the concentration of long polymers is maintained by autocatalysis and how this affects the selected population structures. Our simple model uses a system of binary polymers, where each polymer can replicate itself by exact ligation of two matching subsequences in addition to a random ligation and decomposition. Copy error mutations in the replication process could be considered with ease, but are left out of the model not to obfuscate our analysis.
We give analytical solutions for the polymer population as well as the result of stochastic simulations, where the concentrations of most polymers are the same as that of monomers if autocatalysis works perfectly. Furthermore, we show that the autocatalysis spontaneously organizes the network into characteristic selected cooperative structures with associated selected patterns of the involved species. This ordering is qualitatively different from the known ordering through self-organized criticality [21]; it originates from the symmetry of the network and the dynamics of the reactions. To our knowledge, this type of self-organization has not previously been reported.

Model. — Our binary polymers are modeled as strings of monomers over the alphabet $A = \{0, 1\}$. We denote the set of polymers by $A^*$ and we write $|k|$ to express the length of polymer $k \in A^*$. In the simplest realization for interaction between polymers, three types of reactions are introduced: decomposition of a strand into any two sub-strands with rate $c_0$, random ligation of two strands with rate $c_1$, and autocatalytic ligation with rate $c_2$. Formally

$$l.m \xrightarrow{c_0} l + m,$$

$$l + m \xrightarrow{c_1} l.m,$$

$$l + m + l.m \xrightarrow{c_2} 2\ l.m,$$

for all $l, m \in A^*$, where $l.m$ represents the concatenation of strands $l$ and $m$. Here, we consider only closed systems where material is conserved. However, we implicitly assume an inflow of free energy, which is necessary to activate the molecules involved in ligation and to separate double strands resulting from autocatalytic replication. Our simple replicator model deliberately lacks many important aspects of realistic polymer replication, such as monomer activation [3] and potential mutations [22], in order to isolate the impact of autocatalysis.

It has been observed experimentally, that non-enzymatic replicators suffer from product inhibition [23], where most potential templates are in their inactive double-strand configuration, and thus cannot serve as replication templates. As a consequence, experimental replicators do not follow the exponential growth dynamics of simple autocatalysis and, in turn, promote coexistence of replicator species, rather than “selection of the fittest” [24]. In our study, the implicit energy flow is assumed to be partly used to separate the inactive and low energy double-strand configuration of replicator strands and to transform them into their activated single-strand configuration — thereby reconstituting exponential growth.

Arbitrary sized, finite populations are represented by a multidimensional vector $x$ where each species $k \in A^*$ defines one dimension and its concentration $x_k \in \mathbb{R}^+$ specifies a coordinate. Assuming the law of mass action, the reaction kinetic equations can be written in terms of the molar concentrations $x_k$ of all species $k$ as

$$\frac{dx_k}{dt} = c_0 \left( \sum_{i,j} x_i - \sum_{i} x_k \right),$$

$$+ c_1 \left( \sum_{i,j} x_i x_j - \sum_{j} x_j x_k \right),$$

$$+ c_2 \left( \sum_{i,j} x_i x_j x_k - \sum_{j} x_i x_j x_k \right) \equiv f_k,$$

where $i, j, k \in A^*$. Summations $\sum_{i,j} \text{ and } \sum_{k,j} \text{ are considered every pair of polymers } i \text{ and } j \text{ which satisfy } i = k, \text{ and } j = i \text{ or } j = k, \text{ respectively. Note that the latter in principle involves an infinite number of summands, but is finite under our assumption of a finite material pool.}$

In the limit of purely random ligation ($c_1 > 0, c_2 = 0$), there is a known exponential stationary solution [4,5,25] that satisfies $f(x^*) = 0$: $x_k^* = (c_0/c_1)e^{-b|k|}$, (5)

where $b$ is a constant determined by the boundary condition. In purely random systems ($c_2 = 0$), there is no selection: species decrease in concentration exponentially proportional to their length $|k|$, but their abundance $x_k$ is otherwise independent of the particular sequence $k \in A^*$. The exponential distribution is not a stationary solution in the presence of exact autocatalysis as assumed in reaction (3).

Population structures. — Every existing species creates via decomposition a set of all its substrings. We call any species that is not a substring of any other present species a “chief” and we call the set of its substrings its “clan”. The notion of the chief-clan structure is useful in the autocatalytic case since a few chief species determine the structure of the entire population, as we shall show below.

Although prebiotically less relevant, it is informative to first consider the purely autocatalytic limit ($c_1 = 0, c_2 > 0$). The autocatalytic systems with spontaneous random ligations will be discussed later on. In this limit, there is a somewhat counterintuitive constant solution

$$x_k^* = \begin{cases} \sqrt{(c_0/c_2)}, & k \in A^1 \setminus \partial A^1, \\ 0, & k \not\in A^1, \end{cases}$$

where $A^1$ is the union of all clans and $\partial A^1$ the set of all chiefs, both given by the initial condition. Interestingly, chief populations $x_m^* \in \partial A^1$ are unconstrained in

28004-p2
Structure and selection in an autocatalytic binary polymer model

![Diagram of chief-clan structures](image)

Fig. 1: (Color online) Structures produced by simulation and averaged over a cluster obtained by HCA analysis. Four most frequent structures (a)-(d) are named: (a) bootlace, (b) pinecone1, (c) pinecone2, and (d) two towers. An example of asymmetric structures is shown in (e), which was obtained from the eighth cluster. The darkness of the color shows the concentration of each molecule, where the darker the color, the larger the concentration. Species having less concentration than 10% of the constant solution in (6) are omitted. Only edges connecting species with two adjacent lengths are drawn.

Since \( x_m^* \neq 0 \), \( \sum_{i,j=m} x_i^* x_j^* = (c_0/c_2)(|m| - 1) \), and \( x_m^* \) is arbitrary, determined only by the boundary condition. Because of mass conservation, the maximal length of chiefs is determined by the initial condition and the reaction rates through eq. (6). In the absence of random ligation, the chief-clan structure is entirely determined by the initial condition. Note that the constant solution hinges on the exact replication without overhangs. In models that include sequence overhang, the exponential distribution can be reestablished, but with significantly longer stable sequences [25].

Examples of chief-clan structures are shown in fig. 1, where polymers maintain each other by ligation and decomposition. Two or more chiefs can coexist as seen, for example, in fig. 1(a), where 01010101010 and 10101010101 coexist and share most of their clan members. The maximal length of the chief, \(|m|\), scales as \(|m| \sim \sqrt{n/x} \), when \( n \gg x \) with \( n \) being the total amount of material.

The usefulness of the chief-clan concept is lost in the case without autocatalysis though they also exist in this case by definition but only with exponentially small concentrations. Therefore, we use the concept only for the structures maintained by autocatalysis.

**Stability analysis.** — We analyze the stability of chief-clan structures in the purely autocatalytic limit (\( c_1 = 0 \)). The entries of the functional matrix \( \mathbf{D} f = \frac{\partial f}{\partial x_i} \) are given by

\[
\frac{\partial f_k}{\partial x_l} = c_0 \left( \sum_{k,j,l} 1 - \delta_{k,l} (|k| - 1) \right) + c_2 \left( \sum_{k,j,l} \delta_{k,l} x_i x_j - \sum_{k,j,l} x_i x_j \right) + x_k \left( \sum_{k,j,l} x_i - \sum_{k,j,l} x_j - \sum_{k,j,l} x_j \right),
\]

where \( \delta_{i,j} = 1 \) if \( i = j \) else 0 is the Kronecker symbol. When the system is in a stationary state \( x^* \) that satisfies eq. (6), and is equilibrated enough so that the population contains all possible substrings of existing strings, the equation simplifies to

\[
\text{Case } k = l: \quad \frac{\partial f_l}{\partial x_l} |_{x^*} = -c_0 \left( \sum_{l,k,j,l} 1 + \sum_{l,k,j,l} 2 \right),
\]

\[
\text{Case } k \neq l: \quad \frac{\partial f_k}{\partial x_l} |_{x^*} = c_0 \left( \sum_{l,k,j,l} 1 - \sum_{l,k,j,l} 1 \right).
\]
Notably, the stability of a stationary population is entirely determined by the topology of the reaction graph, not by the reaction rates. $c_0$ merely sets the “speed” with which a stable state is approached but does not determine its absolute stability.

Equations (9) and (10) are expected from detailed balance: increasing the concentration of one reactant, will increase product concentration and decrease the concentration of other reactants with rates given by the equations. The equations also confirm that if $l = m$ is a chief, $\partial f_k/\partial x_m = 0$ for all $k$—chiefs are unconstrained in any stationary stable state.

To evaluate the linear stability of chief-clan structures, the characteristic polynomial, $|Df - \lambda I| = 0$, was solved numerically to give $\lambda_i$ for a multitude of trial populations: i) all 511 chief-clan structures generated by a single chief of length up to eight; ii) chief-clan structures generated by a randomly selected single chief of length up to 20. With numerical noise assumed to be smaller than $10^{-8}$, all of the population have eigenvalues with non-positive real parts, which strongly suggests that stationary populations with a single chief are locally stable in general, independent of their length. In addition, we calculated the eigenvalue spectra of iii) all possible populations with several chiefs with length up to four; and iv) 10000 randomly generated populations with four chiefs of lengths up to 18. We find that the majority of these mixed populations is locally stable, but the probability to encounter unstable structures tends to increase with the length and number of chiefs. For example, about 8% of a chief-clan structure with two chiefs of length 12 is unstable.

Summarizing, stability analysis indicates that there are numerous locally stable chief-clan structures. Yet, we will see below that only very few chief-clan structures out of the vast number of possible structures are selected by the reaction dynamics.

Selection of structures. – We now consider the prebiotically interesting case where ligation is mostly autocatalytic but with a small rate of random ligation ($0 < c_1 \ll c_2 \bar{x}$). Random ligations can disturb otherwise stable chief-clan structures, and introduce novel chiefs that compete with existing chiefs for shared resources or completely absorb the existing chiefs into their clan. It will be seen that the constant solution with unconstrained chiefs drives the selection of few chiefs with highly regular sequence patterns.

In this case the dynamics are strongly dependent on numerical fluctuations, and we thus sample stochastic trajectories from numerical simulation of the discrete reaction system (1)–(3), using the exact stochastic simulation algorithm by Gillespie [26]. The system is assumed to be well stirred and its volume is set to 1. Simulations are started in the state of 200000 monomers of 0 and 200000 monomers of 1. The chemical reaction constants are set to $c_0 = 1$, $c_1 = 1 \times 10^{-10}$, and $c_2 = 1 \times 10^{-7}$, unless otherwise noted. With $c_0 = 1$ the unit of time $t$ is the relaxation time of decomposition.

Figure 2(a) shows how $x_k$ obtained by simulation depends on the length $|k|$. The exponential solution is always observed in purely random systems ($c_2 = 0$, solid circles in fig. 2(a)) as predicted. The constant solution, $\bar{x} = \sqrt{c_0/c_2}$, on the other hand, is always observed in purely autocatalytic systems ($c_1 = 0$, fig. 2(a)). With the parameters chosen, selected chiefs have the length equal to 11. The concentration of the chiefs is also confirmed to be free from $\bar{x}$, as a point deviated from $\bar{x}$ in fig. 2(a). The dependence of $\bar{x}$ on $c_2$ is shown in fig. 2(b). The theory ($\bar{x} = \sqrt{c_0/c_2}$, solid line) predicts exactly the values obtained by the simulation. This constant solution is independent of the initial condition.

In contrast to the random ligation scenario, autocatalytic systems with small rate of random ligation ($c_1 \ll c_2 \bar{x}$) generate only very few distinctive and highly ordered population structures and their sequence patterns are selected out of a huge number of possibilities. We classify the populations obtained by simulation with a single linkage hierarchical cluster algorithm (HCA) [27] based on cosine distance $d(x, x') = 1 - \cos(x/|x|, x'/|x'|)$ for two populations $x$ and $x'$. Clustering is stopped when the number of clusters becomes less than 20% of all the populations.

Using 1200 populations produced by simulation ($t = 100$), HCA shows that only four structures occupy about 67% of all the populations produced. We call them, from the most frequent one, bootlace, pinecone1, pinecone2, and two towers (fig. 1). The probabilities of their appearance at $t = 100$ are about 46%, 14%, 5%, and 2%, respectively. Note that populations with these distinctly regular structures always have a Jacobian with non-positive real eigenvalues. The fifth (2%), sixth (1%), and seventh (1%) clusters are transient structures that...
Structure and selection in an autocatalytic binary polymer model

Fig. 3: (Color online) A stabilization process of “bootlace” structure. The darker the color the larger the concentration. (a) $t = 1.1$ (b) $t = 1.2$ (c) $t = 1.3$ (d) $t = 2.1$ (e) $t = 4.8$ (f) $t = 230$. Only edges connecting species with two adjacent lengths are drawn.

eventually turn into pinecone1. An asymmetric structure appears at last as the eighth cluster (1%), where the dominant chief is 11100100 as shown in fig. 1(e).

It can be seen that the four selected structures have only few regular chiefs (lighter or darker species in fig. 1) and maximal clan overlap. Though fluctuations sometimes create shorter chiefs, their concentration is always much lower than $x$, and they usually disappear quickly.

To explain the mechanisms of this selection process, a typical system evolution is shown in fig. 3. After spontaneous formation by random ligation, an intermediate chief (10) accumulates material through autocatalysis (fig. 3(a)) in order to satisfy the constant solution (eqs. (6) and (7)). Thus, the population is “inverted” and the next random ligation most likely occurs among these chiefs. Once created, chiefs decompose to produce their clan members (fig. 3(b)$\rightarrow$(c) and (d)$\rightarrow$(e)). This continues until chief concentrations become lower than $x$ at which point there is not enough material to stabilize longer chiefs. The described process creates a regularity of the chiefs; e.g., an intermediate chief 10 creates another intermediate chief 1010, which then creates 10101010. Irregular chiefs, on the other hand, require the formation of additional intermediate sequences by random ligation, which is significantly less likely as the population has moved most material into existing chiefs. This sequence pattern selection spontaneously appears out of the autocatalytic ligation-decomposition chemistry in the systems.

Fig. 4: The temporal change of cosine similarity $I(t)$ between the structure at $t = 0$ and the one at $t$. (a) Bootlace, solid line and two towers, dotted line. (b) Pinecone1, solid line and pinecone2, dotted line. Note the scale difference of the y-axis.

The overlap of clans introduces resource competition. Survival against this competition is biased, since the material flow rate from the existing species to new chiefs is not equal among the clans: a chief with more and larger populated exclusive clan members is expected to collect material faster and is therefore more likely to survive competition. Thus a short chief is likely to disappear. Also, the selection process is stronger the larger the system, thus making random ligations of non-chief members relatively rarer.

The selection mechanism is absent in purely random ligation because the exponential equilibrium solution (eq. (5)) does not allow population structures where material is accumulated in long species. There is no driving force to push the system into creating longer species. Thus the selection only arises out of autocatalysis.

We now discuss the stability of the four selected structures against fluctuations. Fluctuations are well known to have a destabilizing effect on intermittently stable structures [21]. It has not been analyzed before, however, that different structures can respond differently to noise. The effect of fluctuations is quantified by the temporal change in cosine similarity $I(t) = \cos(x(0)/|x(0)|, x(t)/|x(t)|)$ between the structure at time $t$ and the initial one. Simulations are started from already stabilized structures.

Bootlace and two towers are mostly stable and exhibit only small fluctuations (fig. 4(a)) compared with those of pinecones (fig. 4(b)). Note the different y-axis scales. The bootlace structure fluctuates between two stable states (10101, 00110, 10010, 00011) as its chief, where the structure suddenly destabilizes after a period of stability, and switches into another stable structure. Another such punctuated equilibrium is observed in pinecone1 (fig. 4(b)), where the structure is switching among the four chiefs, 0011, 0110, 1100, and 1001.

Even a transition from one structure to a completely different one can occur. Pinecone2 (fig. 1(c)) is metastable under this boundary condition, and after a long time, suddenly transitions into a different structure (sudden drop of the dotted line in fig. 4(b)).

Conclusions. – Our model introduces the simplest and most symmetric possible ligation-decomposition chemistry
into a binary polymer system, where every species can serve as product, educt as well as catalyst. This idealization allows us to treat the reaction kinetics of the system analytically. Despite its simplicity, random variations are introduced into the population through the random ligation and decomposition reactions, while the polymer selection and amplification processes occur through template-directed ligation. These reactions together drive the population through a transient evolutionary process to a stable structure satisfying the boundary conditions. Summarizing the main findings for the model: firstly, we find that the constant solution (6) appears out of exact autocatalysis and forces the population to a chief-clan structure, which is in most cases linearly stable. This finding is confirmed by stochastic (Gillespie) simulation. The appearance of the chief-clan structure explains why and how long polymers survive in template-directed reaction systems studied here and, e.g., by Derr et al. [25]. Secondly, when starting from a pool of monomers, only few sequences are selected by the dynamics. Selection does not require an a priori fitness function. This aspect was not elucidated by previous studies where selection was either induced by means of an external fitness function [1] or distinct, randomly attributed catalysis rates [14,15]. Template-directed autocatalysis always plays an important role in the theory of the origin of life, where even specific sequences of primordial molecular replicators have been suggested [6]. Yet, to our knowledge, it has not been reported previously that the selection of specific sequence patterns arises spontaneously out of the autocatalytic dynamics. This intrinsic selection is important for the study of the origin of complex and functional polymers, since it is questionable whether symmetric polymers such as 010101... observed in our simulation happen to carry prebiotic functionality.

These findings are independent of reaction rates as long as $c_1 \ll c_2$. They are also robust against a number of obvious model extensions. The introduction of low-level copy-error mutations or mass flow, larger alphabet size, or sequence complementarity does not change the essence of our findings, while naturally it does increase the complexity of the phenomena. In contrast to our result, Derr et al. suggest that autocatalysis enhances the diversity of polymer species [25]. The discrepancy between their finding and ours probably originates from the potential sequence overhangs incorporated in their model. This issue is currently subject to close investigation.

The current study does not address how our strongly autocatalytic systems arise [4]. It also needs to be clarified more systematically how more realistic and complex mechanisms, such as synthesis of monomers [5], explicit activation reactions [4], mutation, deletion and insertion [22], strand complementarity, sequence overhang [25], and hybridization behavior [5,19,20] impact the structures and their selection in our model. However, we believe that our simple model provides a starting point for elucidating these more realistic situations which eventually could provide clues about expected early information patterns in the origin of life.

HF received financial support from the European Commission under grant agreement CM1304. SR is supported in part by the European Commission MICREA Agents project.

REFERENCES

[1] EIGEN M., Naturwissenschaften, 58 (1971) 465.
[2] COLGATE S. A. and ZIOCH H., Complexity, 16 (2011) 54.
[3] MONNARD P.-A., in Prebiotic Evolution and Astrobiology, edited by WONG J. T. and LAZCANO A. (Landes Bioscience) 2008.
[4] Wu M. and HIGGS P. G., J. Mol. Evol., 69 (2009) 541.
[5] Wu M. and HIGGS P. G., Astrobiology, 11 (2011) 895.
[6] EIGEN M. and WINKLER-OSWALLT R., Naturwissenschaften, 65 (1981) 282.
[7] Von KIEDROWSKI G., Angew. Chem., 25 (1986) 932.
[8] Joyce G. F., Nature, 335 (1989) 217.
[9] KANAVARIOTI A., J. Mol. Evol., 46 (1998) 622.
[10] ROHATGI R., BARTEL D. P. and SZOSTAK J. W., J. Am. Chem. Soc., 118 (1996) 3340.
[11] Sawai H. and WADA M., Orig. Life Evol. Biosph., 30 (2000) 503.
[12] KAUFFMAN S. A., J. Cybern., 1 (1971) 71.
[13] RASMUSSESON, PhD Thesis, Technical University of Denmark, Lyngby (1985).
[14] Farmer J. D., KAUFFMAN S. A. and Packard N. H., Physica D, 22 (1986) 50.
[15] KAUFFMAN S. A., J. Theor. Biol., 119 (1986) 1.
[16] RASMUSSESON S., MOSEKILDE E. and ENGELBRECHT J., in Structure, Coherence and Chaos in Dynamical Systems, edited by CHRISTIANSEN P. L. and PARMENTIER R. D. (Manchester University Press, Manchester) 1989.
[17] Bagley R. J. and Farmer J. D., in Artificial Life II, edited by Langton C. G., Taylor C., Farmer J. D. and Rasmussen S. (Addison-Wesley) 1991. p. 93.
[18] Hordijk W., KAuffmann S. A. and STEEL M., Int. J. Mol. Sci., 12 (2011) 3085.
[19] Fernando C., Von KIEDROWSKI G. and STEEL M., Int. J. Mol. Sci., 64 (2007) 572.
[20] FELLERMANN H. and RASMUSSESON, Entropy, 13 (2011) 1882.
[21] AwaZu A. and KANEKO K., Phys. Rev. E, 76 (2007) 041915; 80 (2009) 010902(R).
[22] Saakian D. B., Phys. Rev. E, 78 (2008) 061920.
[23] Kiedrowski G., Włotzka B., Helbing J., Matzen M. and Jordan S., Angew. Chem., Int. Ed., 30 (1991) 423.
[24] Szathmáry E. and Gladkii I., J. Theor. Biol., 138 (1991) 55.
[25] Derr J., Manapat M. L., Rajamani S., Leu K., Xulvi-Brunet R., Joseph I., Nowak M. A. and Chen I. A., Nucleic Acids Res., 40 (2012) 4711.
[26] Gillespie D. T., J. Comput. Phys., 22 (1976) 403.
[27] Johnson S. C., Psychometrika, 32 (1967) 241.

28004-p6