Sequence selection and evolution in a binary polymer model

Steen Rasmussen[1,2], Harold Fellermann[1], Shinpei Tanaka[3]

[1]Center for Fundamental Living Technology, University of Southern Denmark, Campusvej 55, 50230 Odense, Denmark
[2]Santa Fe Institute, Santa Fe, 1399 Hyde Park Rd, New Mexico, USA
[3]Graduate School of Integrated Arts and Sciences, Hiroshima University, 1-7-1 Kagamiyama, Higashi-Hiroshima 739-8521, Japan
steen@sdu.dk, harold@sdu.dk, shinpei@hiroshima-u.ac.jp

The emergence of autocatalytic structures in model chemistries has been a prominent subject throughout the history of artificial life research (Rasmussen (1985); Farmer et al. (1986); Kauffman (1986); Rasmussen (1989); Hordijk et al. (2011) and others). Most of these works have been concerned with the likelihood of finding autocatalytic sets in a population of random cross-catalytic molecules. Here, in contrast, we study how the detailed sequence structure determines the properties of the emergent cooperative structures. In particular, we study a system of binary polymers, where each polymer can replicate itself by exact ligation of two matching subsequences. We report the emergence of stable cooperative structures with high equilibrated polymer concentrations together with a quantitative connection between the details of the sequence and the frequency as well as the stability of the evolving cooperative structures. These findings could have implications for early earth information polymers as well as the design of protocell information polymer networks.

In the simplest realization, we allow for decomposition, random ligation, and autocatalytic ligation of polymers via the three reactions

$$l.m \xrightarrow{c_0} l + m$$
$$l + m \xrightarrow{c_1} l.m$$
$$l + m + l.m \xrightarrow{c_2} 2\, l.m,$$
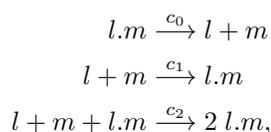
where l and m are strings of arbitrary length over the alphabet $\{0, 1\}$ and $l.m$ denotes string concatenation. c_0, c_1, and c_2 are the respective reaction rates.

If random ligation is comparably rare ($c_1 \ll c_2\bar{x}$ for some typical species size $\bar{x}$), the system exposes several unexpected dynamics: (1) out of exponentially many possible strings, stochastic simulation (Gillespie (1977)) repeatedly selects only very few specific strands; (2) the selected populations are strikingly regular — with the motif 010101 being most common; (3) the occupancy of most strings in a population fluctuates around a constant value independent of the length of the string (c.f. Fig. 1) All these properties are in direct contrast to a scenario without catalysis ($c_2 = 0$).

Most of these features can be explained analytically. Examining the reaction kinetic equations reveals that every string of a stationary state must be accompanied by its substrings which are generated through decomposition. We call the longest members of a population its "chiefs" and the sets of their substrings "clans". In the limit $c_1 \to 0$, chiefs can have an arbitrary occupancy, whereas all non-chief members of a population equilibrate to a constant value determined by the rates of autocatalysis and decomposition as $\bar{x} = \sqrt{c_0/c_2}$. Linear stability analysis of several hundred exemplary stationary states confirms that these populations are indeed stable states. Under purely random ligation ($c_2 = 0$), on the other hand, the stationary strand distribution is exponential in strand length k: $x_k = (c_0/c_1)e^{-b|k|}$, where b is given by the boundary condition.

When starting from a pool of monomers, the ligation dynamics will transport most of the material from the monomers to the emerging front of a chief-clan structure. Rare random ligations slowly expand this front by forming new chiefs. To explain why regular chiefs, such as 01010101, are selected more frequently than irregular ones we note: (i) regular chiefs require the formation of fewer intermediate chiefs: formation of 01010101, for example, requires three random ligations, whereas formation of 01101100 requires at least five ligations; (ii) regular chiefs offer more reaction pathways than irregular ones, as substrings of irregular strings have to ligate in correct order; (iii) population size strongly impacts the reaction rates which results in selection of certain regular sequence patterns.

Thoroughly calculating the likelihoods of forming given chief structures from monomer pools by occasional random ligation and subsequent equilibration, confirms quantitatively that this is indeed the driving mechanism of selection in the simulation results. These calculations also suggest that the "twotowers" structure (Fig. 1.d) becomes less likely with increasing system size, whereas system size shows little impact on the probabilities to obtain other regular structures.

While autocatalysis and degradation stabilize clan structures, random ligation introduces fluctuations in the sys-
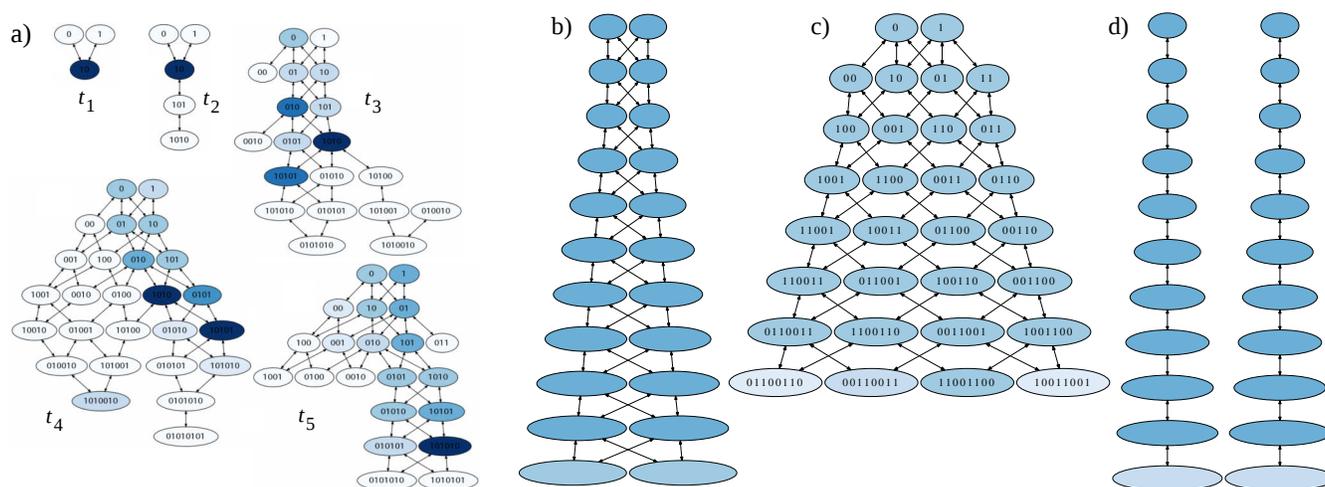
Figure 1: Representatives of the three most prominent stable cooperative structures (b)-(d) found in stochastic simulation at $t = 100$, together with snapshots (a) of the evolutionary formation of structure (b). More than 4000 simulations with parameters $c_0 = 1$, $c_1 = 10^{-10}$, $c_2 = 10^{-7}$ and initial condition $x_0 = x_1 = 100,000$ are performed. Graphs (b)-(d) show the average of all respective members of the three biggest clusters. Results have been aggregated using single-linkage hierarchical clustering using cosine distance. Nodes represent strings with molecular occupancy greater than or equal to 100, where darker color means larger population. Links among non-adjacent layers are omitted for clarity. 44% of the simulations resulted in a bootlace structure generated by the chiefs containing the motif 10101010 (b); another 19% of the simulations produced a pinecone structure with chiefs containing the motif 00110011 (b); and another 2% of the simulations generated the two-tower structure (d) with chiefs 00000000 and 11111111. The shown structures cover more than 65% of the simulations.
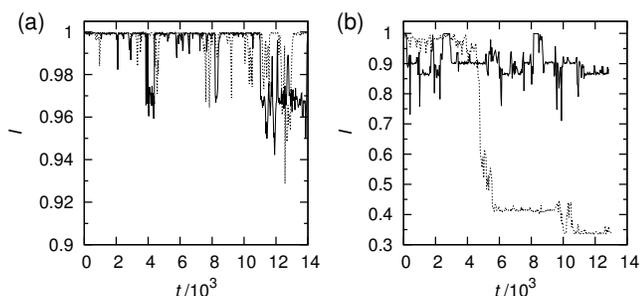

Figure 2: Fluctuations seen in the cosine similarity between an initial structure at $t = 0$ and the one at t for (a) bootlace (solid line) and twotowers (dotted line), and (b) different pinecone structures. Note different y-axis scales.

tem. Fig. 2 shows fluctuations observed in populations drawn from the most prominent selected clan structures. While "bootlace" and "twotowers" structures (Fig 1.b and d) are relatively robust against fluctuations and reestablish the initial population after perturbations, "pinecone" structures (Fig 1.c) are more susceptible to randomness and the populations move between several metastable states which correspond to the four possible competing chiefs seen in Fig 1.c. This demonstrates how random events through selection generate punctuated equilibria in the evolutionary dynamics.

The dynamics and evolutionary potential of these systems with open boundary conditions (chemostat dynamics) remain to be studied, as well as the impact of a chemically more realistic, but also more complex, template complementarity based matching with possible sequence overhang.

Acknowledgements

References

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

Gillespie, D. T. (1977). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, 22(4):403–434.

Hordijk, W., Kauffman, S. A., and Steel, M. (2011). Required levels of catalysis for emergence of autocatalytic sets in models of chemical reaction systems. *Int J Mol Sci*, 12(5):3085–101.

Kauffman, S. A. (1986). Autocatalytic sets of proteins. *J. Theor. Biol.*, 119:1–24.

Rasmussen, S. (1985). PhD thesis, Technical University of Denmark, Lyngby.

Rasmussen, S. (1989). Toward a quantitative theory of the origin of life. In Langton, C., editor, *Artificial Life*, pages 79–104. Addison-Wesley.