

Autocatalytic sets of self-replicating RNAs

01.10.09

Motivation and Background

The origin of life (OoL) on earth is still in search of a satisfactory explanation. The field is dominated by many facets and partial explanations. Facets include frequency of planetary systems, climatology of early earth, chirality in naturally occurring compounds, abiotic production of molecules, etc. Partial explanations include the naturally occurring self-reproducing molecules, quasispecies and hypercycles, an RNA world, natural formation of micelles, etc. However, most of these explanations are incomplete or based on speculation.

Many other fields in the biosciences have benefited from the introduction of formal models, which forced researchers to be explicit about assumptions made, and allowed mathematical reasoning to be applied and computational experiments to be performed. Such models have been introduced in the context of the OoL, but research in them has not been very dominant so far. However, as OoL research gains pace, they will be given more attention. Examples of formal models related to OoL are Conway, von Neumann (1967), Ganti (1997), Kauffman (1986), and Steel (2000). For formal models to be useful they should capture some essence of the empirical problem and as time passes they should be forced towards increasingly realistic descriptions of the phenomena. The formalisation of catalytic reaction systems by Steel (2000), based on an initial idea by Kauffman, consists of

- a set of molecule types;
- a set of reactions where each reaction converts one set of molecules (reactants) into another set (products);
- a set of catalysations: molecules that accelerate a reaction (or set of reactions);
- a food set: a small set of molecules assumed to be freely available and constantly replenished.

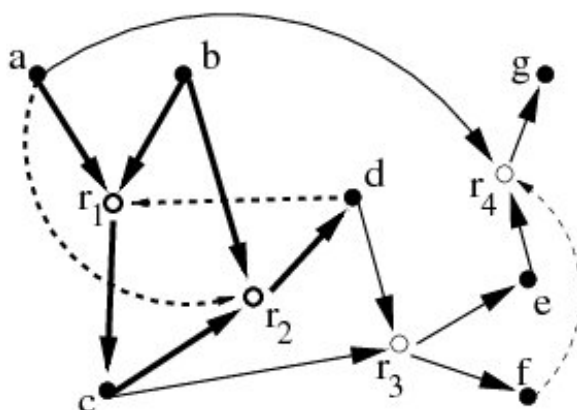


Figure 1: An example of a catalytic reaction system (CRS). The subset of reactions $\{r_1, r_2\}$ (shown in bold) is an RAF set.

The questions of interest in these models are conditions for the appearance of (sub)sets of molecules/reactions that are self-sustainable: each reaction in the set is catalysed by at least one molecule from the set, and each molecule can be created, starting from the food set, by repeated reactions from the same set. This idea of *autocatalytic sets* was introduced in Kauffman (1986), and formalised as *RAF sets* and subsequently studied more extensively in Steel (2000), Hordijk and Steel (2004), and Mossel and Steel (2005). Investigations into these models represent significant progress relative to less precise models, but the Steel model needs elaboration to be more realistic in addressing the probability of spontaneous occurrences of RAF sets.

Project proposal

The current project considers self-sustaining and self-replicating RAF sets of molecules and proposes an extension towards more realism. In particular, we want to investigate two particular (related) properties:

- *RNA-like molecule types.* In the previous studies on RAF sets, molecule types were represented by bit strings. We want to extend this to RNA-like polymers consisting of 4 bases (A, C, G, and U), with base-pair complementarity (A-U and C-G). We will keep the ligation and cleavage reactions of the original studies, i.e., two RNA polymers can be “glued” together into a larger one (e.g., **GAC+AUCA**→**GACAUCA**), or one RNA polymer can be split up into two smaller ones (e.g., **ACGUGA**→**AC+GUGA**).

- *Catalysis based on template matching.* In the previous studies, catalysis was assigned at random according to some (fixed) probability. In an arbitrary mixture of molecules and reactions, it is not always known which molecules can catalyse which reactions, and such a random approach would be justified. However, we want to investigate the specific case where catalysis is driven by the base-pair complementarity of RNA molecules. For example, in the ligation reaction $\text{GAC} + \text{AUCA} \rightarrow \text{GACAUCA}$, we could allow as possible catalysts other RNA molecules of the type **GU** and **UGUA**, i.e., those molecules that match (in terms of base-pair complementarity) the reactants around the ligation site. And similarly for cleavage reactions. Figure 2 below illustrates the idea of such template-based catalysis.

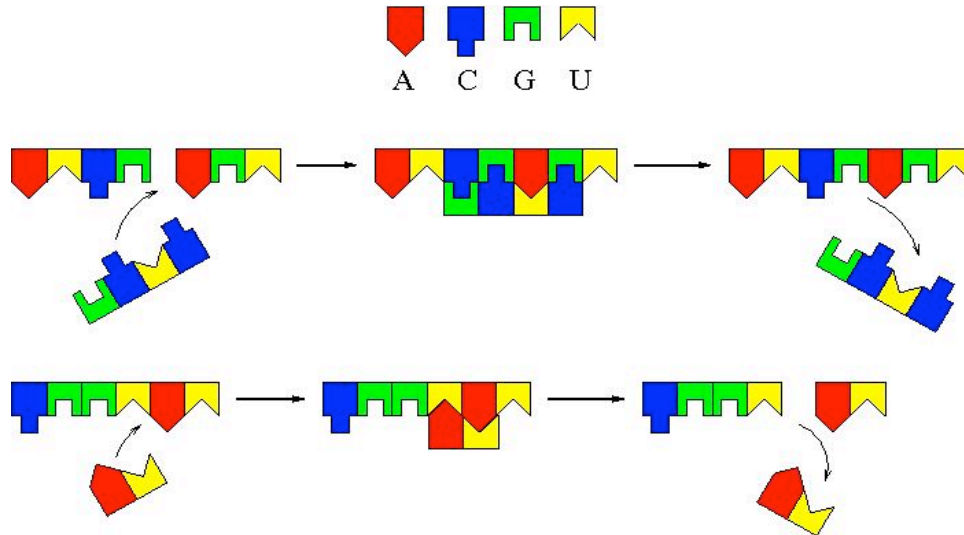


Figure 2: An example of reactions with template-based catalysis. The first reaction is the ligation of two short RNA molecules into one longer one. The second reaction is the cleavage of one longer RNA molecules into two shorter ones.

This extended model with RNA-like polymers and catalysis driven by base-pair complementarity will be analysed both theoretically and computationally. Similar to the previous studies on RAF sets, we will try to derive upper bounds on the probability of RAF sets occurring in this more realistic setting. Furthermore, we will perform computer simulations of these template-based RNA reaction networks and analyse them in terms of frequency of occurrence and size distribution of RAF sets. This work can be based on an already existing implementation of the RAF framework (written in C), which will be modified for the current proposal. Crucial issues to be investigated are

- How does the resulting RAFs depend on food set? The food set must contain dinucleotides otherwise no ligation is possible. But if the foodset only contains a few longer molecules, the RAFs will most likely be a very small subset of all possible RNAs.
- How does the resulting RAFs depend on the stringency of complementarity need to catalyze? Less stringency will clearly create larger RAF sets. (in this context it would interesting to investigate continuous analogues of these Boolean RAFs, where concentration can be described.)

Plan:

- i. Read the background literature.
- ii. Write simulator that can take foodset and iteratively construct longer RNAs.
- iii. Implement the Hordijk-Steel algorithm to identify RAFs.
- iv. Investigate the obtain RAFs as function of foodset and catalysis rules
- v. continuously write report

References

- Ganti. Biogenesis itself. *Journal of Theoretical Biology* 187, pp. 583-593, 1997.
- Hordijk and Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology* 227(4), pp. 451-461, 2004.
- Kauffman. Autocatalytic sets of proteins. *Journal of Theoretical Biology* 119, pp. 1-24, 1986.
- Mossel and Steel. Random biochemical networks: the probability of self-sustaining autocatalysis. *Journal of Theoretical Biology* 233(3), pp. 327-336, 2005.
- Steel. The emergence of a self-catalysing structure in abstract origin-of-life models. *Applied Mathematics Letters* 3, pp.91-95, 2000.
- Von Neumann. *Self-reproducing automata*. 1967.