

Search for Life in Catalytic Reaction Systems

Jotun Hein & Rune Lyngsø

June 25, 2010

Background

The origin of life 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 von Neumann [1966], Gardner [1970], Ganti [1997], Kauffman [1986], 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 (CRS) by Steel [2000], based on an initial idea by Kauffman [1986], 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.

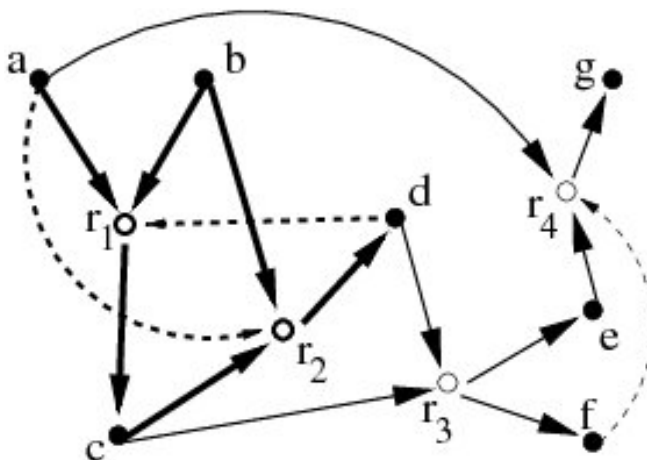


Figure 1: An example [Hordijk and Steel, 2004, Fig. 1] of a catalytic reaction system. The subset of reactions $\{r_1, r_2\}$ (shown in bold) is a reflexive autocatalytic F-generated (RAF) set for $F = \{a, b\}$

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], 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.

Suggested Project Work

In Mossel and Steel [2005] Theorem 4.1 and Corollary 4.2 only apply to the random sequence-based model, introduced in Kauffman [1986], where reactions correspond to ligation and cleavage of sequences. At the very origin of life, it is unlikely that the sequence nature observed in present day life had yet been fully introduced, so it would be interesting to generalise the analysis to apply to arbitrary sets of molecules and reactions, either taken from a database as discussed

in Hordijk et al. [2010] or randomly generated according to some parameters. Another alternative generalisation would be to assume that molecules are represented by sequences of chemical groups, but where reactions would transfer a coherent group, i.e. a substring, between two reactants – in this setup ligation would be a special case where the entirety of one molecule is transferred to the beginning or end of the other molecule. Catalysis remains random, and subject to some simple model as e.g. described in [Mossel and Steel, 2005, (R1) & R2]. The maths problem would be to give upper/lower bounds similar to [Mossel and Steel, 2005, The. 4.1 & Cor. 4.2] or efficient algorithms for computing upper and lower bounds for specific systems.

Alternatively, one could take a step towards the life side of origin of life, and constrain catalysation to only occur when the catalyst is complementary (or matches) the ligation/cleavage site of a reaction. This is described in more detail in Hein [2010]. Assuming that all matching templates for a ligation/cleavage will catalyse the corresponding reaction results in a fully determined system once sequence lengths and required template overlaps have been specified. In most cases this will also result in systems where RAFs are trivially either present or not present. So interesting challenges would be defining realistic scenarios for random catalysis for which we can prove results similar to [Mossel and Steel, 2005, The. 4.1 & Cor. 4.2]. The nice thing is that in all cases we can simulate systems according to the chosen model and use the algorithm in Hordijk and Steel [2004] to efficiently determine the presence of RAFs, as this algorithm makes no assumptions about the structure of the CRS. This can be very useful in developing initial conjectures, and to test the tightness of bounds obtained.

The polynomial time algorithm presented in Hordijk and Steel [2004] for detecting the presence of a RAF in a CRS $\mathcal{Q} = (X, \mathcal{R}, C)$, and for finding a minimal RAF is based on two observations. For a set of reactions \mathcal{R}' removing a reaction $r \in \mathcal{R}'$ that is not catalysed by an element $x \in \text{supp}(\mathcal{R}')$ will not change the maximal RA of the system. Similarly, removing a molecule $x \in X$ that cannot be generated from the food set will not change the closure of the food set relative to \mathcal{R}' either. So we can iteratively remove such reactions and molecules until no more exist, and this will generate the maximal RAF. If this is non-empty, we can proceed to tentatively remove remaining reactions until the resulting RAF becomes empty, at which point we have a minimal RAF.

There are several natural ways to attempt to improve this algorithmic framework. One possibility is to improve on the efficiency of the algorithm to find maximal and/or minimal RAFs. In Hordijk and Steel [2004] the analysis of the presented algorithm yields a worst case complexity of $O(|X||\mathcal{R}|^3)$ while empirical results indicate an average run time of $O(|\mathcal{R}|^{1.43})$. The algorithm essentially starts from scratch in each iteration when identifying reactions or molecules that

should be removed. Reactions are removed when there are no more molecules catalysing them and molecules are removed when there are no more reactions producing them, so it is plausible to assume that an algorithm based on reference counting can achieve linear or close to linear running time. This would significantly expand the size of systems that can be analysed – given the above empirical average running time and the size of $5 \cdot 10^6$ of largest systems analysed mentioned in Hordijk et al. [2010], the size of manageable systems would increase by approximately three orders of magnitude – and should be pursued. Though these algorithms for identifying maximal RAFs can easily be modified to find one minimal RAF, they are less well suited for the problem of counting or enumerating all minimal RAFs. It should be possible to enumerate all minimal RAFs by an exhaustive elimination of each reaction in turn from larger RAFs and maintaining the set of minimal RAFs identified so far, but this will be a rather time consuming approach. Developing methods for faster enumeration and counting, possibly approximately, of minimal RAFs will make it much more feasible to analyse the density and competitiveness in large CRS.

When adding molecule inhibition of reactions to the model, the problem of finding RAFs is proven **NP** complete in Mossel and Steel [2005]. This does not necessarily end the story, though. It is possible to develop heuristic approaches that finds a solution for many instances of the problem. A very naïve approach is simply to run the RAF algorithm ignoring inhibitions – if no RAF is found then there is no RAF when inhibitions are included, and if a RAF is found where none of the F -generated molecules inhibit any of the reactions then this proves the existence of a RAF. One can expand on this idea or develop novel approaches that will allow the question of the presence of RAFs to be settled for most random systems. As long as the set of systems where no decision is reached is small, this will still allow a simulation based investigation into the likelihood of the emergence of RAFs under different parameter values.

A different approach to developing heuristic methods is to identify conditions under which it is no longer **NP** hard to find RAFs in the presence of inhibitions. In Mossel and Steel [2005] no assumptions were made about the nature of inhibitions. However, one can explore the effect of several natural restrictions of inhibitions. For example, inhibition of reaction r by molecule x could be caused by x binding to one of the reactant molecules a of r . In such a case it would be natural to require that all reactions where a is a reactant are inhibited by x . Alternatively, the inhibition could be a simple matter of kinetics. If r and r' are two reactions utilising the same reactant molecule x but r running much faster than r' , then inhibition could be a case of r depleting x . Hence it would be more a case of r inhibiting r' , i.e. molecules could only inhibit ‘neighbouring’ reactions.

References

- Tibor Ganti. Biogenesis itself. *Journal of Theoretical Biology*, 187(4):583–593, 1997.
- Martin Gardner. The fantastic combinations of John Conway’s new solitaire game “life”. *Scientific American*, 223(120–123), 1970.
- Jotun Hein. Autocatalytic sets of self-replicating RNAs. http://www.stats.ox.ac.uk/__data/assets/pdf_file/0004/5377/Autocatalytic_sets_of_RNAs.pdf, 2010.
- Wim Hordijk and Mike Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227(4):451–461, 2004.
- Wim Hordijk, Jotun Hein, and Mike Steel. Autocatalytic sets and the origin of life. *Entropy*, 2010. Submitted.
- Sanjay Jain and Sandeep Krishna. Autocatalytic sets and the growth of complexity in an evolutionary model. *Physical Review Letters*, 81(25):5684–5687, 1998.
- Stuart A. Kauffman. Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119(1):1–24, 1986.
- William Martin and Michael J. Russell. On the origin of biochemistry at an alkaline hydrothermal vent. *Philosophical Transactions of the Royal Society B*, 362(1486):1887–1926, 2007.
- Elchanan Mossel and Mike Steel. Random biochemical networks: the probability of self-sustaining autocatalysis. *Journal of Theoretical Biology*, 233(3):327–336, 2005.
- David Penny. An interpretive review of the origin of life research. *Biology & Philosophy*, 20(4):633–671, 2005.
- Mike Steel. The emergence of a self-catalysing structure in abstract origin-of-life models. *Applied Mathematics Letters*, 13(3):91–95, 2000.
- John von Neumann. *The Theory of Self-Reproducing Automata*. 1966.