

Computational Models for Origin of Life

Objective: To give a presentation 60 minutes at the end of the course and write 20+ page report covering the key aspects of how to computationally model the origin of life also discussing which research could be done.

The origin of life (OoL) remains an elusive problem, although much progress has been made in recent years. Having gone from a problem considered beyond science to solvable in principle but maybe not in practice, many researchers now believe it will be solved in the next few decades. In fact, there are currently several OoL scenarios, but all still have some difficulties and lacunae to be overcome. The occurrence of an RNA world preceding DNA and proteins is generally considered to have been an important step in the origin of life. However, an RNA world spontaneously generating complex proteins and DNA still poses some unresolved problems. Moreover, and perhaps more importantly, the appearance of the RNA world itself is still mostly an open question. It has been argued that a simple prebiotic metabolism could have generated the basic building blocks of life. This would have involved an autocatalytic cycle (such as the reverse citric acid cycle), i.e., a self-reinforcing cyclic arrangement of chemicals and reactions, each element responsible for producing the next one in the cycle. Although this idea is still debated, there is experimental evidence that simple (auto)catalytic networks could indeed have produced the basic ingredients for an RNA world. The appearance of proteins and then DNA (possibly from a precursor RNA world) is supported by arguments using ideas such as hypercycles, Darwin-Eigen cycles, or collectively autocatalytic sets. The claimed spontaneous appearance of autocatalytic sets has been disputed, but there is some evidence that they could be constructed experimentally. Summarizing the above story, a plausible scenario for the origin of life that has recently emerged is as follows: (Inorganic chemistry → Prebiotic metabolism → RNA world → Proteins → DNA. Although none of these steps have (yet) been proven beyond any doubt, they are at least plausible, and there is evidence (theoretical and/or experimental) to support them.

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. This project tries to discuss different formal models for the origin of life.

The field is presently experiencing a major rise in importance and funding. There are several reasons for this:

- The rise of synthetic biology, where increasingly large biosystems are being designed *de novo* which paves the way for Origin of Life experiments
- Within the solar systems probes are being sent to the major candidates where life could exist – Mars, Europe and Titan. Intelligent tests for life cannot be designed without proper definitions of “life”
- The very large number extra-solar planets that now have been found – including earth-sized ones make prevalence of life in the Universe increasingly likely.

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.

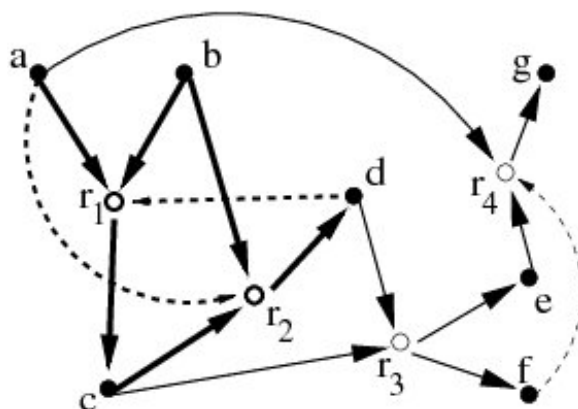


Figure 1: An example of a catalytic reaction system (CRS) from Hordijk and Steel (2004). *a, b* are here food. The subset of reactions $\{r_1, r_2\}$ (shown in bold) is an RAF set.

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. Wim, Hein and Hordijk (2010) sketches some possibilities.

The Big Questions Are: How could the origin of life problem be solved? What are the main present challenges in solving this? What can be learnt from models of origin of life? Which properties must such models have?

Possible Contents of Presentation/report

The History of Origins of Life Research

History of Models in Origin of Life Research

The Key Models:

Von Neumann model

Eigen: Quasi-species and Hypercycles

The Chemoton

Kauffman-Steel Model

The Key Challenges in Computational Origin Research

Possible starting points

Artificial Life vol 4.3 (1998) Special Issue on Self Replicating Automata.

Berlekamp, Conway and Guy (1982) *Winning Ways I-II* chapt. 25 Academic Press.

Eigen, "Self-organization of Matter and the Evolution of Biological Macromolecules," *Naturwissenschaften* 58 (1971): 465-523

Eigen M. and P. Schuster. *The Hypercycle: A principle of natural self-organization*, Springer, Berlin, 1979

Faulon, Visco and Roe (2005) "Enumerating Molecules" *Rev. Compu. Chem.*, Vol.21

Ganti, T (1997) *Biogenesis itself*. *Journal of Theoretical Biology* 187, pp. 583-593, 1997.

Harary and Palmer (1973) *Graphical Enumeration* Acad. Press.

Hordijk and Steel (2004) Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J. Theor Biology* 227(4), pp. 451-461,

Hordijk, Hein and Steel (2010) Autocatalytic Sets and the Origin of Life *Entropy* 2010, 12, 1733-1742

Kauffman (1986) Autocatalytic sets of proteins. *Journal of Theoretical Biology* 119, pp. 1-24, 1986.

Meringer (2010) "Structure Enumeration and Sampling" in *Handbook of Chemoinformatics Algorithms* CRC Press

Mossel and Steel (2005) Random biochemical networks: the probability of self-sustaining autocatalysis. *J. Theor Biology* 233(3), pp. 327-336,

Maynard Smith, J & E. Szathmari (1995) "Major Transitions in Evolution." Chaps.1-7

Sloane and Nambi (2006) "Integer Sequences Related to Chemistry"

Randi (2004) "Nenad Trinajstić – Pioneer of Chemical Graph Theory" *CROATICA CHEMICA ACTA* 77 (1-2) 1-15

Steel (2000) The emergence of a self-catalysing structure in abstract origin-of-life models. *Applied Mathematics Letters* 3, pp.91-95,

von Neumann, John (1966). A. Burks. ed. *The Theory of Self-reproducing Automata*. Urbana, IL: Univ. of Illinois Press

Possible starting points

<http://www.liebertpub.com/products/product.aspx?pid=99>

<http://www.springer.com/life+sciences/journal/11084>

http://www.ismb.lon.ac.uk/origin_of_life.html