

Evolution Without Smart Molecules

Chris Gordon-Smith

Independent Researcher
c.gordonsmith@gmail.com
<http://www.simsoup.info>

Abstract. Network dynamics may have played a key role in the Origin of Life. ‘Smart’ molecules such as template replicators and enzymes may not have been necessary in the first evolving entities. Instead, a process of natural selection between chemical networks operating in different organisms may have been the key evolutionary mechanism. This paper shows such a process using the SimSoup artificial chemistry simulation. The context and conceptual background for SimSoup is first outlined. The model is then described, and differences with other models are highlighted. SimSoup has network elements that correspond directly to the unimolecular and bimolecular elementary reaction schemes of physical chemistry. These network elements can be combined in very general ways to produce ‘compound interactions’ which can be catalytic. The model includes mass conservation, reaction rates based on considerations of energy and thermodynamics, and cycle detection. A run of the model is presented showing an evolutionary process in which a metabolic network is modified at periodic intervals, and the modification is accepted or discarded according to whether or not it results in higher metabolic activity. The network includes a large number of cyclic flows. It evolves through a series of persistent states, each of which can be regarded as a different ‘species’.

1 Introduction

1.1 Origin Of Life Issues and Views

Solving the problem of the Origin of Life entails finding an explanation for the Origin of Evolution. Since Darwin, the origin of new species has been well understood. By contrast, there is no complete and generally accepted explanation for the Origin of Evolution. Two key questions are:-

1. How did entities capable of transferring inherited information arise?
2. What were these entities made of?

There are two main views on the first question:-

- **Genetic View:** Template replicating molecules or crystals were crucial for the Origin of Life, and have from the outset been the carriers of the inherited information that makes evolution possible
- **Metabolic View:** The first living entities were metabolic systems, and they evolved by exploring the possibilities for new kinds of metabolic network.

On the second question, the following views can be identified:-

- **Heterotrophic View:** The first living entities constructed and maintained themselves using preformed organic molecules synthesised in a ‘primordial soup’ by non-life chemical processes
- **Autotrophic View:** The first living entities constructed and maintained themselves using small readily available inorganic molecules.

Much attention has focused on the RNA World and other genetic theories (including Cairns-Smith [1]), but a full explanation for the origin of accurately replicating template molecules remains elusive. The main problem is the extreme improbability of such molecules arising through a purely random process.

Metabolic View theories take a simpler starting point. In their heterotrophic variants they require sufficient concentrations of preformed organic molecules such as amino acids or lipids. In autotrophic variants the starting point is even simpler; small inorganic molecules are all that is required. Appealing to simplicity does not of course relieve Metabolic View theories of the obligation to explain how a metabolic system with no template replicator can evolve. Finding such an explanation is the motivation for this paper.

1.2 Smart Molecules or Networks?

Since the discovery of the structure of DNA there has been a focus on the role of template replicating molecules in biology. The insights and achievements that have followed are real and very substantial. There is however a growing sense that life cannot be explained solely in terms of complex molecules with highly specialised properties.¹ I refer to template replicating molecules, and the complex enzymes that are required for accurate replication, as **‘smart’ molecules**.

The SimSoup project seeks to investigate the role of non-genetic mechanisms in the Origin of Life. In particular, SimSoup is about understanding the dynamic properties of chemical networks, and the role these could have had in the first evolving systems. The aim is not to show that a chemical network can produce a ‘smart’ molecule that can carry inherited information and evolve; it is to show that a chemical network containing only ‘dumb’ molecules has sufficient information carrying properties to enable evolution to begin.

SimSoup does not make any assumption about whether the prebiotic world was heterotrophic or autotrophic. The network oriented evolutionary process envisaged in this paper is applicable to either case.²

1.3 Conceptual Background for SimSoup

The conceptual background for SimSoup includes the following:-

- The Metabolic View theories of Aleksandr Oparin (Oparin [8]), Stuart Kauffman (Kauffman [5]), Freeman Dyson (Dyson [2]), and more specifically the Lipid World and the GARD model of Doron Lancet’s group (Segré et al. [9] [10] [11])
- Network theory, particularly the work of Sanjay Jain and Sandeep Krishna (Jain and Krishna [4], Krishna [6]).
- Günter Wächtershäuser’s chemo-autotrophic Iron-Sulphur World (Wächtershäuser [13] [14] [15])

¹ The popular misconception that the DNA code ‘causes’ life has been called ‘DNA-mania’ by André Pichot (referenced in Noble [7]).

² The name ‘SimSoup’ does not imply an *organic* soup. “You can have a soup of anything” (Wächtershäuser, quoted at <http://www.simsoup.info>).

GARD and The Lipid World: It is well known that amphiphiles, a type of lipid molecule, can spontaneously form *vesicles* (also known as *liposomes*); small spherical structures in which a lipid bilayer membrane encloses an aqueous compartment. The Lipid World envisages growth of the membrane due to the production of amphiphiles and their incorporation into the membrane. Organic ‘food’ molecules are used to generate the amphiphiles in reactions that are catalysed by amphiphiles already present in the membrane. When a vesicle reaches a critical size, it splits giving rise to offspring.

The rate of growth of the membrane depends on the overall rate of production of amphiphiles, and will be high when the membrane includes an autocatalytic set of amphiphiles that mutually catalyse each other’s formation through a network of catalytic relationships.

When a vesicle divides, its offspring have almost the same composition (relative frequencies) of amphiphile species as the parent. As a result, the offspring usually inherit the catalytic properties of the parent, and will have a similar growth rate. Occasionally variations occur, and a new population of vesicles with different characteristics is originated. A process of natural selection begins in which the fastest growing population is the one whose members have the most effective autocatalytic network.

The Lipid World is a heterotrophic theory; it assumes the presence of pre-formed organic molecules from which the amphiphiles can be constructed.

The GARD model is based on The Lipid World scenario. It envisages a set of lipid species denoted L_i, L_j, L_k etc. The catalytic network is represented by a matrix β , such that β_{ij} represents the enhancement to the rate of production of L_i due to the presence of L_j . The ‘food’ molecules are assumed to have constant concentration, and numerical integration is used to examine the effect of mutual catalysis on the time dependent concentrations of the lipids.

A Model Based on Graph Theory and Matrix Algebra: Jain and Krishna [4] have used the techniques of graph theory to model metabolic networks. Each molecular species is represented as a node in a directed graph with nodes labelled $i \in S \equiv \{1, 2, 3, \dots, s\}$ and with adjacency matrix $C \equiv (c_{ij})$. If there is a directed link from species j to species i then $c_{ij} = 1$, otherwise $c_{ij} = 0$. The relative populations of the different species are represented by a vector \mathbf{x} , with x_i being the relative population of species (node) i . A directed link between two nodes represents the catalytic influence of one species on the production of another. The directed link $c_{ij} = 1$ means that j is a catalyst for the production of i . The dynamic behaviour of \mathbf{x} is modelled according to the rate equations for catalysed reactions in a well stirred reactor, with the simplifying assumption that the non-catalytic ‘food’ reactants for the reactions have constant concentration.

Krishna [6] identifies some potential drawbacks of this assumption. However, a strength of the problem formulation is that it enables the powerful techniques of graph theory and matrix algebra to be used to draw a number of conclusions. These include: i) \mathbf{x} is the eigenvector of C with the largest eigenvalue λ_1 , ii) if $\lambda_1 \geq 1$, then the set of nodes for which $x_i > 0$ constitutes the ‘dominant

autocatalytic set’, iii) the dominant autocatalytic set includes a ‘core’ and a ‘periphery’, with nodes in the core contributing to each other’s maintenance, whereas nodes in the periphery are parasitic, and iv) λ_1 is a measure of the core size and the multiplicity of pathways or ‘redundancy’ within it.

Jain and Krishna analyse a scenario in which the network C is changed periodically such that a species with low population is removed (along with its links), and a new node is added with random links to and from other nodes. This leads to crashes in which several species are rapidly depopulated, followed by recoveries. It is shown that the crashes typically occur when $\lambda_1 \approx 1$.

The Iron-Sulphur World: Wächtershäuser has proposed an autotrophic ‘pioneer organism’ with a mineral substructure and an organic superstructure. Iron and sulphur are key constituents of the substructure. The organic compounds on the superstructure arise by reductive carbon fixation (redox reactions) involving molecules such as CO, CO₂, H₂S, N₂, and HCN. These reactions are catalysed by metal centres in or on the surface of the sub-structure. As the organic compounds are produced they bond to the substructure, so that the organism grows. Some of the organic molecules catalyse the reactions that produced them. This positive feedback effect results in an autocatalytic network of reactions; Wächtershäuser [14] notes that this means reproduction. The autocatalytic feedback effect is subject to variation, which he identifies as the basis for evolution.

The primary focus of Wächtershäuser’s work is on the specific reactions relevant for the pioneer organism, and he analyses candidate reactions in detail.

2 The SimSoup Model

2.1 Comparison with Other Models

SimSoup is an abstract model of a chemical network and its dynamic behaviour. Before describing the model, some comparisons with GARD and the model of Jain and Krishna will help to set the context:-

1. The basic element of a SimSoup network corresponds to an elementary (unimolecular or bimolecular) reaction, not to a catalytic relationship. There is no analogue of the β and C matrices of the GARD and Jain/Krishna models
2. Catalytic behaviour is not ‘built in’ to SimSoup. Instead, different Interaction Types can combine in ways that correspond to catalysis
3. There is no assumption of constant concentration for Reactants. The concentrations of both Reactants and Products vary with time
4. Each Molecule Type has a mass, and mass is conserved
5. Energy and thermodynamics determine Interaction rates
6. Each individual Molecule of each type (species) exists as a separate object in the simulation.³ This enables behaviour to be monitored at a detailed level. This includes the detection of cycles.

³ SimSoup is implemented as an object-oriented computer program. When an Interaction occurs, Reactant Molecule object(s) are destroyed and Product Molecule object(s) are created.

Items 1 to 5 make SimSoup a realistic (yet simple and abstract) model of chemical networks. The avoidance of ‘built in’ catalysis means that any autocatalytic behaviour is purely a network property; there is no assumption of ‘smart’ molecules with high specificity. This is particularly relevant in the context of autotrophic scenarios where catalytic effects are typically of low specificity. As a result of item 3; the matrix mathematics used by Jain and Krishna are not in general applicable to SimSoup networks. However, the simulation approach (item 6) enables the modelling of scenarios that would otherwise be mathematically intractable.

2.2 The Basic SimSoup Model

The basic SimSoup model is described in Gordon-Smith [3], and so is covered here in outline only. The laws of Chemistry are represented by a set of **Molecule Types** and **Interaction Types** that set a framework within which dynamic behaviour can take place. A Molecule Type is analogous to a molecular species in real chemistry. It has a mass, and can be associated with Interaction Types in which individual **Molecules** participate as **Reactants** or **Products**. Apart from this, and its potential energy (see below), a Molecule Type has no other properties; it is essentially a network-oriented entity.

An Interaction Type is a category of interaction between Molecule Types. Each Interaction Type has forward and reverse rate constants, k_f and k_r , that are used to determine the overall rate. There are three forms of Interaction Type, representing the elementary reaction schemes of physical chemistry. Table 1 shows these three forms. Elementary reactions are either unimolecular or bimolecular, and the table shows the molecularity for Interactions of each type.

Form of Interaction Type	Interaction Scheme	Interaction Rate
Construction (bimolecular)	$A + B \rightarrow C$	$k_f ab - k_r c$
Transformation (unimolecular)	$D \rightarrow E$	$k_f d - k_r e$
Fission (unimolecular)	$F \rightarrow G + H$	$k_f f - h_r g$

Table 1. The three forms of Interaction Type in SimSoup. k_f and k_r are the forward and reverse rate constants. Other lower case letters (a, b etc.) represent the concentrations of the corresponding (upper case) Molecule Types. Interaction Schemes shown indicate the forward reactions. The molecularity shown also relates to forward reactions.

Interaction Types can be combined in ways that enable **Compound Interactions** to take place. These can have catalytic properties. For example, $A + X \rightarrow I$ followed by $I + B \rightarrow J$ and finally $J \rightarrow X + C$ together make up a Compound Interaction with overall scheme $A + B \xrightarrow{X} C$, with X operating as a catalyst, and I and J as intermediates. Gordon-Smith [3] gives further examples.

Reactions take place in a well-stirred Reactor of constant volume at the rates shown in Table 1.

2.3 Model Extensions

This section describes extensions to SimSoup since Gordon-Smith [3].

The Manhattan Plot: At any time, the SimSoup **Reactor Composition** is defined by the number of Molecules of each type present. This can be expressed as a vector \mathbf{R} . Each element of \mathbf{R} corresponds to the (integer) number of Molecules of a particular type. The dynamic behaviour of the system is defined by a series of (instantaneous) Reactor Compositions, $\mathbf{R}(t_i)$, where t_i is the i_{th} timestep.

\mathbf{R} may be static, or may be moving through composition space (indicating changing numbers of Molecules of the various types). To measure this, the Manhattan Distance $D(\mathbf{R}(t_1), \mathbf{R}(t_2))$ between the Reactor Compositions at times t_1 and t_2 can be calculated follows:-

$$D(\mathbf{R}(t_1), \mathbf{R}(t_2)) = \sum_i |r_{1i} - r_{2i}|.$$

Summation is over Molecule Types. r_{1i}, r_{2i} are the i_{th} elements of $\mathbf{R}(t_1), \mathbf{R}(t_2)$.

SimSoup uses a ‘Manhattan Plot’ to show the scale of differences between Reactor Compositions at different times. Each point in the plot corresponds to a pair of times t_1 and t_2 , and the tone of the plot at that point indicates the Manhattan Distance between $\mathbf{R}(t_1)$ and $\mathbf{R}(t_2)$.

Most unperturbed networks have an attractor in which \mathbf{R} is static. However, in some cases the attractor is cyclic and the Reactor Composition repeatedly traverses a ‘narrow corridor’ in composition space.⁴

Energy, Thermodynamics and Rate Constants in SimSoup: It is found experimentally that many chemical reactions have a rate constant that varies with temperature according to the Arrhenius equation $k = Ae^{-E_a/RT}$, where T is temperature, R is the gas constant, E_a is the activation energy for the reaction, and A is a constant, sometimes known as the ‘frequency factor’. The equation holds for both unimolecular reactions and bimolecular reactions, and for reactions taking place both in gases and in solutions.

SimSoup assigns a potential energy to each Molecule Type and an activation energy to each Interaction Type. Rate constants (forward and reverse) for Interaction Types are then set in accordance with the Arrhenius equation. The heat released (or absorbed) by an Interaction, dQ , is the difference between the potential energy of the Reactant(s) and that of the Product(s). The assignment of a potential energy to each Molecule Type ensures that where Reactant(s) are converted to Product(s) through a series of Compound Interactions, dQ is independent of the specific path taken. This is essential to ensure that SimSoup represents the conservation of energy. Without a fixed potential energy for each Molecule Type, it would be possible for cycles to develop in which energy was not conserved.

⁴ See http://www.simsoup.info/SimSoup/News_2006.html for an example.

Trackers and Cycles: Cycles are key to the operation of autocatalytic sets. In order to enable detection of cycles, SimSoup includes **Trackers**. A Tracker is an object that can be attached to a Molecule. When an Interaction takes place, any Trackers attached to the Reactant(s) are passed on to the Product Molecule(s). If there is more than one Product, the Trackers are distributed at random among them. Trackers are typically attached to a small proportion of Molecules, so that they enable monitoring of a representative sample of the Reactor Interactions.

The path of each Tracker is monitored, enabling completion of cycles to be detected. Each cycle has a **Cycle Type**, which is the sequence of Molecule Types visited. Complete cycles with the same sequence but different starting points are considered to have the same Cycle Type.

3 Evolution of a High Entropy Producing Network

3.1 Model Setup and Results

Figure 1 presents the results of a model run in which a network was evolved with selection for high entropy production. The temperature in the Reactor was constant, and so this also entailed selection for high production of heat (dQ/dt).

The model was setup with 100 Molecule Types, and initially no Interaction Types (no network connections). Five Molecule Types were chosen as ‘food’, and the number of each of these was held constant at 50. To avoid uncontrolled growth of the number of Molecules, a slow leakage was setup; at each timestep each Molecule had a removal probability of 0.0001.

The model was run for 19 million timesteps, with a ‘mutation step’ taking place at intervals of about 1000 timesteps. At each mutation step a Construction or Fission was either added or removed at random. A ‘mutation’ can be regarded as the result of a perturbation to the network, possibly due to a change in the molecular composition when a ‘protocell’ divides, or due to the input from the environment of some novel molecules.⁵ 1000 timesteps after the mutation, if dQ/dt was lower than at the time of the mutation then the mutation was reversed, otherwise it was kept. Reversal of a mutation entailed removing an added Construction/Fission, or reinstating one that had been removed.

3.2 Interpretation

Figure 1b shows that the rate of heat release (dQ/dt) has a general upward trend after time 6.8 million. There are many step change increases and decreases of varying sizes. Some decreases are very large, and can be described as ‘crashes’. In most cases, the effects of a ‘bad’ mutation are reversed by reversing the mutation, but in the case of crashes the damage done by a mutation is irreversible. Crashes are usually followed by later step change recoveries of comparable size.

⁵ See section 4.2 regarding a future enhancement to model such perturbations in a more realistic way.

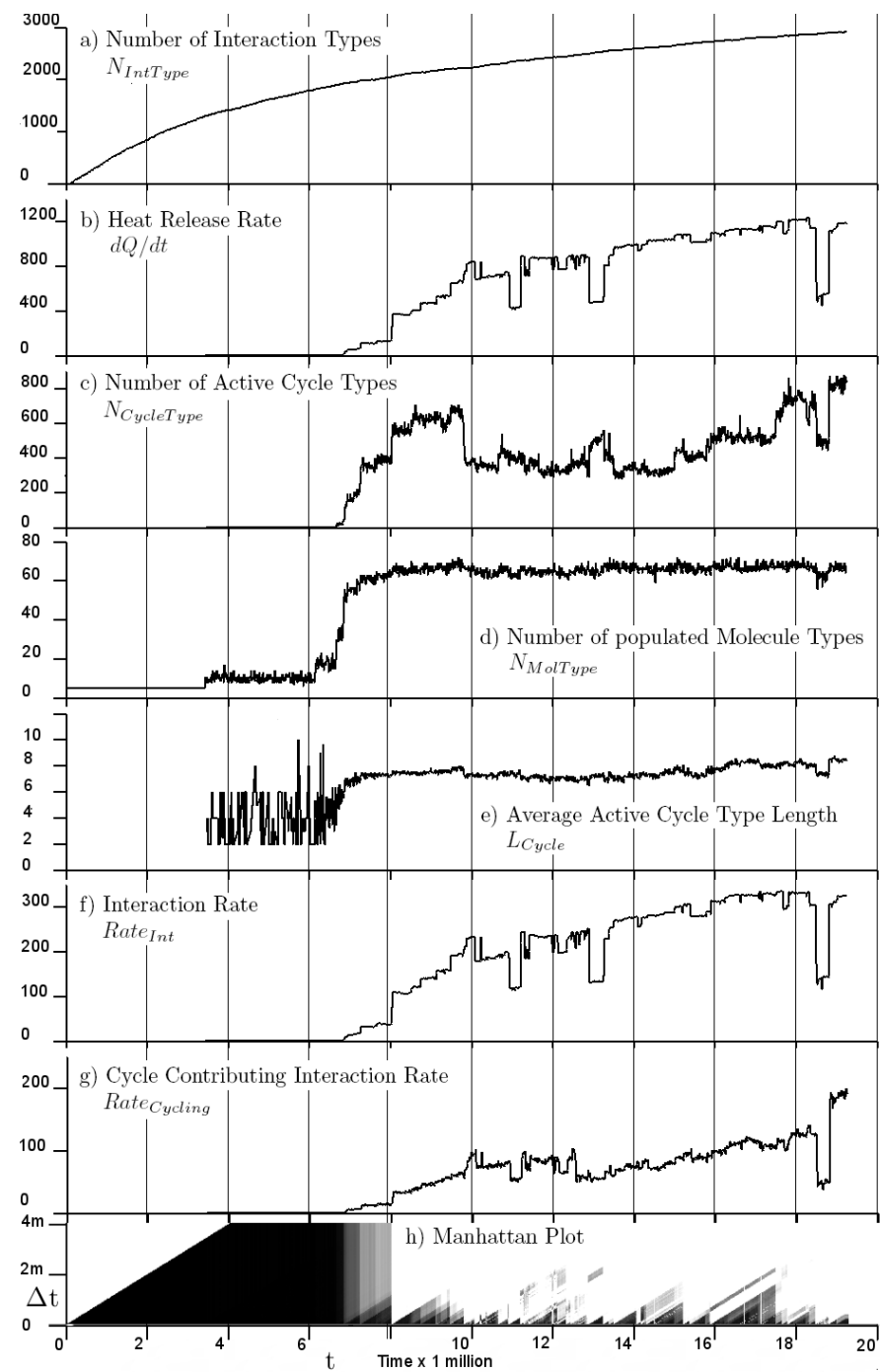


Fig. 1. (a) - (g): Time series plots for the network evolution. (h): Manhattan Plot

The increase in number of Interaction Types shown in figure 1a indicates that selection for high entropy production favours high network connectivity.

Figure 1c shows that the number of Active Cycle Types ($N_{CycleType}$) has step increases and decreases, many of which coincide with the step changes in dQ/dt . It is not the case however that a high dQ/dt always corresponds to high $N_{CycleType}$. At times 9.8 million and 13.4 million an increase in dQ/dt coincides with a decrease in $N_{CycleType}$. From time 13.4 million to 17.5 million $N_{CycleType}$ is lower than at time 8.2 million; the reverse is true for dQ/dt . This suggests that simplifications in network flows can in some cases lead to a higher dQ/dt .

The number of Active Cycle Types is high; at the end of the run it is over 800. It would clearly be incorrect to characterise the overall behaviour as a ‘metabolic cycle’. It is evident that we have a large number of interacting cycles.

Figure 1d shows that $N_{MolType}$, the number of Molecule Types that are represented in the Reactor, also exhibits step change increases. The initial increases coincide with increases in $N_{CycleType}$ (the first increase occurs at a time when $N_{CycleType}$ is very low, and only just visible on the plot).

Figure 1e shows that the average length of active cycles remains at about 8 after an initial variable phase, the end of which coincides with the major increase in $N_{MolType}$.

Figures 1f and 1g show the total Interaction Rate ($Rate_{Int}$) and the rate of Interactions that contribute to Active Cycle Types ($Rate_{Cycling}$). It can be seen that a substantial amount of the activity in the Reactor is cyclic; at the end of the run, 200 of the 320 Interactions occurring at each timestep are contributing to cycles. This indicates that 62.5% of the activity in the network is in the autocatalytic ‘core’.

Figure 1h is the Manhattan Plot for the run. The tone of each point indicates the Manhattan Distance $D(\mathbf{R}(t), \mathbf{R}(t - \Delta t))$ between the Reactor Compositions at time t and a time Δt earlier. Black indicates zero distance. White indicates a distance greater than $\frac{1}{10}D_{max}$, where D_{max} is the maximum possible distance.

The plot shows that before time 6.8 million the Reactor composition \mathbf{R} was almost constant (as indicated by the dark area on the left of the plot). The dark triangles after time 8 million correspond to periods during which $\mathbf{R}(t)$ is approximately constant. The right hand edge of each triangle indicates a sudden change in $\mathbf{R}(t)$. The periods of constant \mathbf{R} last for up to 1.5 million timesteps and represent persistent states of the network. Transitions between persistent states often coincide with step changes in dQ/dt (eg at $t = 8$ million).

4 Conclusions and Prospects

4.1 Conclusions

An abstract model of chemistry has been presented. The model represents chemistry at the level of elementary reactions (unimolecular and bimolecular). This enables chemical networks to be modelled in a very general way, and with catalysis as a property of the network, not of ‘smart’ molecules.

The model conserves mass and energy, and rate constants are determined thermodynamically. No ‘constant concentration’ assumption is made for ‘food’ reactants. Entropy and heat release are measured, and cyclic flows are detected.

High activity networks including a large number of co-existing cycles of different types can develop. ‘Side reactions’ to individual cycles (see Szathmary[12]) do not destroy network activity. It is likely that this is because material leaving one Cycle Type supports others.

Evolving the network for high entropy leads to sharp increases and crashes in network activity. The crashes are similar to those of Jain and Krishna. As the network is evolved it has stable and distinct Reactor Compositions that persist over extended periods, even in the face of regular mutations. Each persistent state of the entire network can be regarded as analogous to a ‘species’. Occasionally a mutation results in a new persistent Reactor Composition. This is similar to the core shifts noted by Jain and Krishna, and can be regarded as analogous to speciation.

In summary, the model demonstrates that a metabolising network with no ‘smart’ molecules or ‘built in’ catalysts can maintain its identity as a ‘species’ and undergo an evolutionary process in which new ‘species’ occasionally arise. The unit of evolution is the entire network, not any particular cycle.

4.2 Prospects

Possible avenues for future investigation include:-

- Addition of new Molecule Types and their Interaction Types in a more open ended and realistic way. Instead of mutations ‘rewiring’ the network of Interaction Types for a fixed set of Molecule Types, new Molecule Types would be constructed in the Reactor (or arrive from the environment), and new Interaction Types would arise naturally as a result of the appearance of these previously unseen Molecule Types
- Investigation of whether a more open ended evolution leads to weakening of the autocatalytic core due to parasitic ‘side reactions’, and under what conditions such ‘side reactions’ are avoided. A tendency for larger Molecules to be less reactive and therefore not undergo Constructions is a possibility (because this would limit the extent of the chemical network and the scope for ‘side reactions’)
- Further investigation of the network paths, flow rates on different paths, and the reasons for crashes and rapid rises in network activity
- Modification of the evolutionary algorithm to restore Reactor Composition after a ‘bad’ mutation as well as reversing the mutation. This would be likely to avoid crashes.

References

1. Cairns-Smith, A.G.: Genetic takeover. Cambridge University Press (1982)
2. Dyson, F.: Origin Of Life. Cambridge University Press (1999)

3. Gordon-Smith, C.: SimSoup: An Artificial Chemistry Model for Investigation of the Evolution of Metabolic Networks. ECAL 2005 Artificial Chemistry Workshop. <http://www.simsoup.info/Publications.html>
4. Jain, S., Krishna, S.: Autocatalytic Sets and the Growth of Complexity in an Evolutionary Model. <http://arxiv.org/abs/adap-org/9809003>
5. Kauffman, S.A.: The Origins Of Order. Oxford University Press (1993)
6. Krishna, S.: Formation And Destruction of Autocatalytic Sets in an Evolving Network Model. PhD Thesis, Indian Institute Of Science (2003), <http://arxiv.org/abs/nlin.AO/0403050>
7. Noble, N.: The Music of Life Oxford University Press (2006)
8. Oparin, A.I.: The Origin Of Life On The Earth. Oliver And Boyd (1957)
9. Segré, D., Dafna, B., Deamer, D., Lancet, D.: The Lipid World. *Origins Life Evol. Biosphere* **31** (2001) 119-145
10. Segré, D., Lancet, D., Kedem, O., Pilpel, Y.: Graded Autocatalysis Replication Domain (GARD): kinetic analysis of self-replication in mutually catalytic sets. *Origins Life Evol. Biosphere* **28** (1998) 501-514
11. Segré, D., Shenhav, B., Kafri, R., Lancet, D.: The Molecular Roots of Compositional Inheritance *J. Theor. Biol.* **213** (2001) 481-491
12. Szathmáry, E.: The origin of replicators and reproducers From volcanic origins of chemoautotrophic life to Bacteria, Archaea and Eukarya *Phil. Trans. R. Soc.* (2006) **361** 1761-1776.
13. Wächtershäuser, G.: Evolution of the first metabolic cycles *Proc. Natl. Acad. Sci. USA* **87** (1990) 200-204
14. Wächtershäuser, G.: From volcanic origins of chemoautotrophic life to Bacteria, Archaea and Eukarya *Phil. Trans. R. Soc.* (2006) **361** 1787-1808.
15. Wächtershäuser, G.: The Origin of Life and its Methodological Challenge. *J. Theor. Biol.* **187** (1997) 483-494