ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Theoretical Biology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Journal of Theoretical Biology

www.elsevier.com/locate/yjtbi

Random biochemical networks: the probability of self-sustaining autocatalysis

Elchanan Mossel[a], Mike Steel[b,*]

[a]*Statistics Department, UC Berkeley, CA, USA*

[b]*Department of Mathematics and Statistics, Biomathematics Research Centre, University of Canterbury, Private Bag 4800, Christchurch, New Zealand*

Received 23 June 2004; received in revised form 23 September 2004; accepted 8 October 2004

Abstract

We determine conditions under which a random biochemical system is likely to contain a subsystem that is both autocatalytic and able to survive on some ambient 'food' source. Such systems have previously been investigated for their relevance to origin-of-life models. In this paper we extend earlier work, by finding precisely the order of catalysation required for the emergence of such self-sustaining autocatalytic networks. This answers questions raised in earlier papers, yet also allows for a more general class of models. We also show that a recently described polynomial-time algorithm for determining whether a catalytic reaction system contains an autocatalytic, self-sustaining subsystem is unlikely to adapt to allow inhibitory catalysation—in this case we show that the associated decision problem is NP-complete.

Keywords: Combinatorial chemistry; Autocatalysis; Discrete random structures

1. Introduction

The idea that the study of discrete random networks could provide some insight into the problem of how primitive life might have emerged from an ambient 'soup' of molecules goes back to mid-1980s. This was largely motivated by the earlier investigation of random graphs, pioneered by Erdös and Rényi in the 1950s and 1960s, which had revealed the widespread occurrence of 'threshold phenomena' (sometimes also called 'phase transitions') in properties of these graphs (Erdös and Rényi, 1960). In the simplest random graph model, one has set of vertices (points) and edges are added independently and randomly between pairs of vertices. As the probability that any two nodes are joined by an edge passes certain well-studied thresholds, there is typically a fundamental change in various qualitative properties of a large random graph, such as its connectivity, or the size of the largest component (see e.g. Bollobas, 2001). Extending this approach, Bollobas and Rasmussen (1989) investigated when a directed cycle would first emerge in a random directed graph, and how many vertices such a cycle would contains. They were motivated by the idea that the emergence of a primitive metabolic cycle was an essential step in the early history of life, writing "we want to know when the first catalytic feedbacks appear, and how many different RNA molecules they involve." Cohen (1988) also foresaw the relevance of random graph techniques for modelling primitive biological processes.

The importance of cycles in early life had also been studied—from a slightly different perspective—by Eigen (1971) and Eigen and Schuster (1979). They proposed a metabolic 'hypercycle' as a way of circumventing the so-called 'error catastrophe' in the formation of longer strings of nucleotides, first demonstrated by Maynard-Smith (1983). The study of such processes and how they

*Corresponding address: Tel.: +64 3 366 7001/7688; fax: 64 3 364 2587.

E-mail addresses: mossel@stat.berkeley.edu (E. Mossel), m.steel@math.canterbury.ac.nz (M. Steel).

0022-5193/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2004.10.011

might further evolve into early life has been extensively investigated, using both stochastic and dynamical approaches (e.g. Scheuring, 2000; Wills and Henderson, 1997; Zintzaras et al., 2002).

The idea that threshold phenomena might help explain some of the mystery surrounding the emergence of life-like systems from a soup of inanimate molecules was developed further by Dyson (1982, 1985) and Kauffman (1986, 1993). Kauffman considered simple autocatalytic protein networks where amino acid sequences catalyse the joining (or 'ligation') of shorter sequences, and the cutting (or 'cleavage') of longer sequences. He calculated that under a simple model of random catalysation, once the collection of sequences became sufficiently extensive there would inevitably emerge a subsystem of reactions that was both autocatalytic and able to be sustained from an ambient supply of short sequences (such as single or pairs of amino acids). Kauffman realized that simple random graphs and digraphs by themselves do not capture the intricacy of chemical reactions and catalysis. A more complex discrete structure—which has become known as a *catalytic reaction system* is required in order to formalize and study the concept of a system of molecules that catalyses all the reactions required for their generation, and which can be sustained from some ambient 'food' source of molecules, F. A different discrete model for self-reproducing systems based on Petri nets has also been developed by Sharov (1991) for investigating the dynamical properties of these systems, but we do not deal with this model here.

Several investigators have developed the study of catalytic reaction systems and random autocatalysis (Hordijk and Fontanari, 2003; Hordijk and Steel, 2004; Lohn et al., 1998; Wills and Henderson, 1997) though it also has its critics (e.g. Lifson, 1997; Orgel, 1992; Maynard Smith and Szathmáry, 1995) and these criticisms are mainly of two types. Firstly, Kauffman invoked overly simplistic and strong assumptions in his analysis—for example he considered just binary sequences (i.e. two amino acids) and assumed that each molecule had the same fixed probability of catalysing any given reaction. In this paper we make much weaker, and thereby hopefully more robust assumptions in our probabilistic analysis. A second concern is more general—the concept of a 'protein-first' start to life is problematic, since proteins, unlike RNA are not able to replicate (for a discussion of this point, part of the so-called 'chicken and egg' problem see Lifson, 1997; Maynard Smith and Szathmáry, 1995 or Penny, 2004). Thus, it is quite likely that other sequences besides proteins (such as RNA) may have been part of the first prebiotic systems, and there has been considerable interest from biochemists in the feasibility of an 'RNA world' in the early stages of the formation of life (for a recent survey, see Penny, 2004).

At this point there are at least two ways to formalize the concept of a self-sustaining and autocatalytic set of molecules—the two we study here are referred to as the RAF (reflectively autocatalytic, and F-generated) and CAF (constructively autocatalytic and F-generated) sets. The former was investigated in Steel (2000) and Hordijk and Steel (2004). A CAF, which we formalize in this paper is a slightly stronger notion—it requires that any molecule m that is involved in any catalysation must already have been built up from catalysed reactions (starting from F). This concept is perhaps overly restrictive, since it might be expected that m would still be present in a random biochemical system in low concentrations initially before reactions that generate a steady supply of m become established.

For the sequence-based models of the type studied by Kauffman, we determine the degree of catalysation required for a RAF or a CAF to arise. In Kauffman's model reactions consist of the concatenation and cutting of sequences up to some maximal (large) length, starting from small sequences of length at most t, and each molecule has a certain probability of (independently) catalysing any given reaction. Let $\mu(x)$ denote the average number of (concatenation) reactions that sequence x catalyses, which may depend on $|x|$ the length of x. Then, roughly speaking, our results show that if $\mu(x)/|x|$ is small the probability that the system contains a RAF is small; conversely if $\mu(x)/|x|$ is large the probability the system contains a RAF is close to 1, and indeed in this case there is likely to be a RAF for which all the molecules in the system are involved. This confirms two conjectures that were posed in Steel (2000) and confirms some trends that were suggested by simulations in Hordijk and Steel (2004).

Our results for RAFs contrast sharply with the degree of catalysation required for a CAF. In that case each molecule needs to catalyse, on average, some fixed proportion of all reactions for a CAF to be likely. That is, the corresponding value of $\mu(x)$ required for a likely occurrence of a CAF is exponentially larger (with n) than for a RAF.

We begin this paper by formalizing the concepts of RAF and CAF, and we do so in a more general setting than Hordijk and Steel (2004) as we consider the effect of general catalysation regimes—for example by allowing certain molecules to inhibit certain reactions. In this case determining whether an arbitrary catalytic reaction system contains a RAF seems to be computationally intractable. Indeed we show that the decision problem is NP-complete. This contrasts with the situation where one allows only positive catalysation; in that case a polynomial-time algorithm (in the size of the system) for finding a RAF if one exists was described in Hordijk and Steel (2004). Sections 4 and 5 present the main results concerning the required growth of $\mu(x)$ with $|x|$ required for RAF and CAR generation, and in Section 6 we

make some concluding comments, and raise some questions for further investigation.

Although the assumptions in Kauffman's original paper were quite strong—for example that each molecule had the same probability of catalysing any given reaction—in this paper we have been able to weaken some of these assumptions. The analysis in this paper still ignores inhibitory catalysis, side reactions that may deplete certain reactants (Szathmáry, 2000), and dynamical aspects of the process (Szathmáry and Maynard Smith, 1995) however, we hope to extend this analysis in future work.

2. Preliminaries and definitions

We mostly follow the notation of Steel (2000) and Hordijk and Steel (2004). Let X denote a set of molecules and $\mathscr{R}$ a set of reactions, where we regard a reaction as an ordered pairs (A, B) where A, B are subsets of X called the *reactants* and *products*, respectively. Let F be a distinguished subset of X, which can be regarded as some plentiful supply ('food') of reactants.

For $r \in \mathscr{R}$ let $\rho(r) = A$ and $\pi(r) = B$ and for a set $\mathscr{R}' \subseteq \mathscr{R}$ let

$$\rho(\mathscr{R}') := \bigcup_{r \in \mathscr{R}'} \rho(r),$$

$$\pi(\mathscr{R}') := \bigcup_{r \in \mathscr{R}'} \pi(r)$$

and

$$\mathrm{supp}(\mathscr{R}') := \rho(\mathscr{R}') \cup \pi(\mathscr{R}').$$

Thus $\mathrm{supp}(\mathscr{R}')$ denotes the molecules in X that are used or produced by at least one reaction in $\mathscr{R}'$.

Given a subset $\mathscr{R}'$ of $\mathscr{R}$ and a subset X' of X the *closure* of X' relative to $\mathscr{R}'$, denoted $cl_{\mathscr{R}'}(X')$ is the (unique) minimal subset W of X that contains X' and that satisfies the following condition for each reaction $(A, B) \in \mathscr{R}'$:

$$A \subseteq X' \cup W \Rightarrow B \subseteq W.$$

It is easily seen that $cl_{\mathscr{R}'}(X')$ is precisely the set of molecules that can be generated starting from X' and repeatedly applying reactions selected (only) from $\mathscr{R}'$.

Let $\gamma : 2^X \times \mathscr{R} \to \{0, 1\}$ be a *catalysation function*. The function γ tells us whether or not each reaction r can proceed in its environment (e.g. be 'catalysed') depending on what other molecules are present. Thus, let $\gamma(A, r) = 1$ precisely when r would be catalysed if the other molecules in the system comprise the set A. For example, consider a simple scenario where each reaction $r \in \mathscr{R}$ is catalysed provided that at least one molecule in some set (specific to r) is present. We can represent the associated function γ as follows—we have a set $C \subseteq X \times R$ (as in Steel, 2000; Hordijk and Steel, 2004) where (x, r) indicates that molecule x catalyses reaction r. The catalysation function $\gamma = \gamma_C$ for this simple setting is then defined by

$$\gamma_C(A, r) = \begin{cases} 1 & \text{if } \exists x \in A : (x, r) \in C, \\ 0 & \text{otherwise.} \end{cases}$$

More generally, suppose we have two arbitrary sets $C(+) \subseteq X \times R$ and $C(-) \subseteq X \times R$, which can represent, respectively, the molecules that catalyse and inhibit the various reactions. Then a candidate for γ is the function $\gamma = \gamma_{C(+),C(-)}$ defined by

$$\gamma_{C(+),C(-)}(A, r) = \begin{cases} 1 & \text{if } \exists x \in A : (x, r) \in C(+) \\ & \quad \text{and there is no } x' \in A : (x', r) \in C(-), \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Thus $\gamma_{C(+),C(-)}$ allows both catalysation and inhibition. We find it useful to write $A \to B$ to denote the reaction (A, B). Similarly, we will write $A \overset{C(+),C(-)}{\longrightarrow} B$ to denote the reaction (A, B) together with a catalysation function that satisfies (1). When the sets A, B, C are singletons we will often omit the { } symbols.

In case γ is monotone in the first coordinate (i.e. $A \subset B \Rightarrow \gamma(A, r) \leqslant \gamma(B, r)$) we will call γ *monotone*. Note that γ_C is monotone, and that monotone catalytic functions do not allow inhibition effects.

The triple $\mathscr{Q} = (X, R, \gamma)$ is called a *catalytic reaction system*.

2.1. Autocatalytic networks

Suppose we are given a catalytic reaction system $\mathscr{Q} = (X, \mathscr{R}, \gamma)$ and a subset F of X.

A *reflexive autocatalytic network over F* or *RAF* for $\mathscr{Q}$ is a non-empty subset $\mathscr{R}'$ of $\mathscr{R}$ for which

(i) $\rho(\mathscr{R}') \subseteq cl_{\mathscr{R}'}(F)$,
(ii) For each $r \in R', \gamma(\mathrm{supp}(\mathscr{R}'), r) = 1$.

In addition, to avoid biological triviality, we will also require that any *RAF* $\mathscr{R}'$ also satisfies the condition

(iii) $\pi(\mathscr{R}') \not\subseteq F$.

Thus for $\mathscr{R}'$ to be a RAF, each molecule involved in $\mathscr{R}'$ must be able to be constructed from F by repeated applications of reactions that lie just in $\mathscr{R}'$ (condition (i)) and each reaction in $\mathscr{R}'$ must be catalysed by the system of molecules involved in $\mathscr{R}$ (condition (ii)). This definition is a slight generalization of that given by Hordijk and Steel (2004) to allow for more general catalysation functions γ in condition (ii). Condition (iii) simply ensures that any set of reactions that produce only molecules that are already in the food set F does not constitute a RAF.

Next, we describe a condition which is somewhat stronger than the RAF requirement.

A *constructively autocatalytic network over F* or *CAF* for $\mathcal{Q}$ is a strictly nested sequence $\emptyset \neq \mathcal{R}_1 \subset \mathcal{R}_2 \subset \cdots \subset \mathcal{R}_k$, for which

(i) $\rho(\mathcal{R}_1) \subseteq F$ and for each $r \in \mathcal{R}_1$, $\gamma(F, r) = 1$.
(ii) For all $i \in \{1, \ldots, k-1\}$, $\rho(\mathcal{R}_{i+1}) \subseteq \text{supp}(\mathcal{R}_i)$, and for each $r \in \mathcal{R}_{i+1}$, $\gamma(\text{supp}(\mathcal{R}_i), r) = 1$.
(iii) $\pi(\mathcal{R}_1) \not\subseteq F$.

Informally, a CAF is a way to sequentially build up a set of molecules, starting with F, and in such a way that every reaction is catalysed by at least one molecule that has already been constructed. Notice that for any catalytic reaction system $\mathcal{Q}$, any set $\mathcal{R}_i$ occurring in a CAF for $\mathcal{Q}$ is also a RAF for $\mathcal{Q}$.

Fig. 1 illustrates these two concepts in the case of simple catalysation (of the form $\gamma = \gamma_C$).

There is a further condition we can impose on a RAF or CAF to make it more biologically relevant—namely we may require that a set of reactions is capable of constructing complex molecules required for maintaining certain biological processes (such as metabolism, error correction or reproduction). Of course, there may be many combinations of complex molecules that suffice to maintain these processes, but we would like $\mathcal{R}'$ to be able to construct at least one of these combinations. We can formalize this notion as follows. Suppose $\mathcal{R}'$ is a RAF (respectively, $\mathcal{R}_1 \subset \mathcal{R}_2 \subset \cdots \subset \mathcal{R}_k = \mathcal{R}'$ a CAF) and suppose $\Omega \subseteq 2^{X-F}$ is a distinguished collection of subsets of molecules. We say that $\mathcal{R}'$ is an (Ω)-*complex RAF* (respectively, an (Ω)-*complex CAF*) if the following condition (ΩC) holds:

(ΩC) If $\Omega \neq \emptyset$ then $C \subseteq \pi(\mathcal{R}')$ for at least one $C \in \Omega$.

We can think of each set $C \in \Omega$ as a suite of complex molecules that are required for maintaining certain biological processes and condition (ΩC) requires that the RAF or CAF be capable of constructing at least one such set. Note that the definition of an Ω-complex RAF (respectively, Ω-complex CAF) reduces to that of a (simple) RAF or CAF if we take $\Omega = \emptyset$.

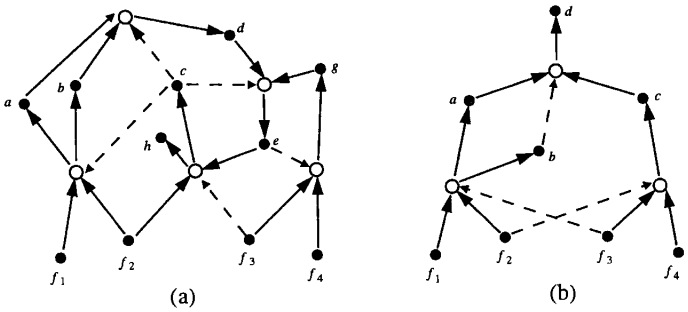

Fig. 1. (a) An example of a RAF and (b) a CAF; represented as directed graphs. Molecules are shown as black nodes, reactions are white nodes, $F = \{f_1, f_2, f_3, f_4\}$ and each (positive) catalysation of a reaction by a molecule is indicated by dashed arc. Solid arcs show the input and output of each reaction.

3. The complexity of determining whether or not $\mathcal{Q}$ has a RAF or a CAF

In Hordijk and Steel (2004) it was shown that, when $\gamma = \gamma_C$, there is a polynomial-time algorithm to determine if $\mathcal{Q}$ has a RAF. However if one allows inhibition also—by replacing γ_C by $\gamma = \gamma_{C(+),C(-)}$—it is unlikely that any efficient algorithm exists for determining a RAF, by virtue of the following result whose proof is given in the Appendix.

Proposition 3.1. *For arbitrary catalytic reaction systems $\mathcal{Q} = (X, \mathcal{R}, \gamma_{C(+),C(-)})$ and a subset F of X the decision problems 'does $\mathcal{Q}$ have a RAF?' is NP-complete.*

However, for any monotone catalysation function γ there is a simple algorithm to determine whether or not $\mathcal{Q}$ has a CAF, which is essentially to let the system 'evolve from F'. We describe this now.

Proposition 3.2. *Given a catalytic reaction system $\mathcal{Q} = (X, \mathcal{R}, \gamma)$, with γ monotone, there is a polynomial time algorithm (in $|X|, |\mathcal{R}|$) to determine whether or not $\mathcal{Q}$ has a CAF.*

Proof. Define a sequence $X_i, \mathcal{R}_i$ for $i \geqslant 1$ as follows:

$X_1 = F; \mathcal{R}_1 = \{r \in \mathcal{R} : \rho(r) \subseteq F, \text{ and } \gamma(F, r) = 1\}$

and for $i \geqslant 1$ set

$X_{i+1} = X_i \cup \pi(\mathcal{R}_i); \mathcal{R}_{i+1} = \mathcal{R}_i \cup \{r \in \mathcal{R} : \rho(r) \subseteq X_i \text{ and } \gamma(X_i, r) = 1\}$.

Then provided $\mathcal{R}_1 \neq \emptyset$ the sequence $\mathcal{R}_1 \subseteq \mathcal{R}_2 \cdots \subseteq \mathcal{R}_k$ (for any $k \geqslant 1$) is a CAF for $\mathcal{Q}$. If $\mathcal{R}_1$ is empty, then clearly $\mathcal{Q}$ has no CAF. □

4. Random sequence-based models

In this section we take $X = X(n)$, the set of sequences of length at most n over the alphabet set $\{0, 1, \ldots, \kappa - 1\}$. Let F be a distinguished (small) subset of $X(n)$; in this paper we will take $F = X(t)$ for a fixed value of t (often a value such as $t = 2$ has been taken in earlier papers). For a sequence $x \in X(n)$ we will let $|x|$ denote its length. Let $\mathcal{R}(n)$ denote the set of all ordered pairs $r = (A, B)$ where, for some $a, b, c \in X$ for which $c = ab$ (= the concatenation of a and b) either $A = \{a, b\}$ and $B = \{ab\}$-in which case we call r a *forward reaction* or $A = \{ab\}$ and $B = \{a, b\}$-in which case we call r a *backward reaction*. We may think of the pair $r = (\{a, b\}, c)$ as representing

the ligation reaction

$a + b \to c$

and the pair $r = (\{c\}, \{a, b\})$ as representing the cleavage reaction

$c \to a + b.$

We will let $\mathscr{R}_+(n)$ and $\mathscr{R}_-(n)$ denote the (partitioning) subsets of $\mathscr{R}(n)$ consisting of the forward and backward reactions, respectively.

Note that we have

$$x_n := |X(n)| = \kappa + \kappa^2 + \cdots + \kappa^n = \frac{\kappa^{n+1} - \kappa}{\kappa - 1}, \tag{2}$$

which is the total number of sequences of length at most n, and

$$r_n := |\mathscr{R}_+(n)| = (\kappa^2 + 2\kappa^3 + \cdots + (n-1)\kappa^n) = \frac{(n-1)\kappa^{n+2} - n\kappa^{n+1} + \kappa^2}{(\kappa - 1)^2}, \tag{3}$$

which is the total number of forward reactions that construct sequences of length at most n. We will often below use the fact that, for all $n \geqslant 1$,

$$1 - O\left(\frac{1}{n}\right) \leqslant \frac{r_n}{nx_n} \leqslant 1 \tag{4}$$

(the notation $f(n) = g(n) + O(\frac{1}{n})$ means $|f(n) - g(n)| \leqslant K/n$ for some constant K for all $n \geqslant 1$).

We study a catalysation function γ obtained by setting $\gamma = \gamma_C$ where C is some random assignment of catalysation (i.e. pairs (x, r)) that is subject to the following requirements:

(R1) The events $((x, r) \in C : x \in X(n), r \in \mathscr{R}_+(n))$ are independent.

(R2) For each sequence $x \in X(n)$ and reaction $r \in \mathscr{R}_+(n)$, the probability $\mathbb{P}[(x, r) \in C]$ depends only on x.

This model is more general than that described in Kauffman (1993), Steel (2000) or Hordijk and Steel (2004) for several reasons—it allows different catalysation probabilities for forward and backward reactions, it allows dependencies involving the catalysation of backward reactions, and the catalysation ability of a molecule can vary according to the molecule considered (for example, it can depend on the length of the molecule).

Let $\mu_n(x)$ be the expected number of reactions in $\mathscr{R}_+(n)$ that molecule x catalyses. By (R2) we can write this as

$$\mu_n(x) = \mathbb{P}[(x, r) \in C] \cdot |\mathscr{R}_+(n)|$$

for any given $r \in \mathscr{R}_+(n)$.

For $\mathscr{Q}(n) = (X(n), \mathscr{R}(n), \gamma_C)$, $F = X(t)$ for some fixed value of t and $\Omega \subseteq 2^{X(n)-F}$, let $\mathscr{P}_n(\Omega)$ be the probability that $\mathscr{Q}(n)$ has an Ω-complex RAF. We can now state the first main result of this paper.

Theorem 4.1. *Consider a random catalytic reaction system $\mathscr{Q}(n)$ satisfying* (R1) *and* (R2) *and with $F = X(t)$ for a fixed value of t, with $t < n$. Let $\lambda \geqslant 0$ and let $\Omega \subseteq 2^{X(n)-F}$.*

(i) *Suppose that $\mu_n(x) \leqslant \lambda n$ for all $x \in X(n)$. Then*

$$\mathscr{P}_n(\Omega) \leqslant 1 - \exp\left(-2\lambda x_t^2\left(1 + O\left(\frac{1}{n}\right)\right)\right) \quad (\to 0 \text{ as } \lambda \to 0),$$

where x_t is defined in (2).

(ii) *Suppose that $\mu_n(x) \geqslant \lambda n$ for all $x \in X(n)$, or that $\mu_n(x) \geqslant \lambda\theta_n|x|$ for all $x \in X(n)$, where $\lambda > \log_e(\kappa)$ and where $\theta_n = \frac{1}{\kappa}(1 + \frac{n\kappa^{n+1}}{r_n}) \sim 1$. Then,*

$$\mathscr{P}_n(\Omega) \geqslant 1 - \frac{\kappa(\kappa e^{-\lambda})^t}{1 - \kappa e^{-\lambda}} \quad (\to 1 \text{ as } \lambda \to \infty).$$

To illustrate Theorem 4.1 consider binary sequences, and a food set consisting of the 6 molecules of length at most 2 (thus $\kappa = t = 2$, which was the default setting for the simulations in Hordijk and Steel, 2004). Then taking $\lambda = 4$ in Theorem 4.1(ii) we have $\mathscr{P}_n > 0.99$.

As an immediate corollary of Theorem 4.1 we obtain the following result, which confirms the two conjectures posed in Steel (2000).

Corollary 4.2. *Consider random catalytic reaction systems $\mathscr{Q}(n)$ $(n \geqslant t)$ satisfying* (R1) *and* (R2) *and with $F = X(t)$ for a fixed value of t. Take $\Omega = \emptyset$, and let $\mathscr{P}_n = \mathscr{P}_n(\emptyset)$, the probability that $\mathscr{Q}(n)$ has a RAF.*

(i) *If*

$$\max_{x \in X(n)} \frac{\mu_n(x)}{n} \to 0 \text{ as } n \to \infty$$

then $\lim_{n\to\infty} \mathscr{P}_n = 0$.

(ii) *If*

$$\min_{x \in X(n)} \frac{\mu_n(x)}{|x|} \to \infty \text{ as } n \to \infty$$

then $\lim_{n\to\infty} \mathscr{P}_n = 1$.

Remark.

- Corollary 4.2 has been worded in such a way that it clearly remains true if we interchange the terms $\frac{\mu_n(x)}{|x|}$ and $\frac{\mu_n(x)}{n}$ in either (i) or (ii) or both.
- The condition described in Corollary 4.2(ii) suffices to guarantee (for large n) a RAF involving all the molecules in $X(n)$. However, it does not guarantee that all of $\mathscr{R}_+(n)$ is an *RAF*. The condition for this latter event to hold with high probability as $n \to \infty$ (assuming for simplicity that $\mu(x)$ is constant,

say μ_n, over $X(n)$) is the stronger condition that

$$\liminf_{n\to\infty} \frac{\mu_n}{n^2} > \log_e(\kappa).$$

This follows from (a slight extension of) Theorem 1 of Steel (2000).

- Note that if we were to view a sequence $(x_1, x_2, \ldots, x_n) \in X(n)$ and its reversal $(x_n, x_{n-1}, \ldots, x_1)$ as equivalent molecules then Corollary 4.2 still holds since asymptotically (with n) palindromic sequences have a negligible influence in the calculations.
- Similarly, if we were to modify (R2) to require that any molecule x cannot catalyse a reaction r for which x is a reactant, then Corollary 4.2 would still hold (and Theorem 4.1 would only be slightly modified) since the number of reactants in any reaction r is asymptotically negligible (with n) compared with the total number of molecules that could catalyse r.
- Note that the lower bound on $\mathscr{P}_n$ in Theorem 4.1(ii) is valid for any value of $n > t$ (previous studies, from Kauffman's (1986) onwards, had drawn conclusions by considering limits as n tended to infinity, but the bound in Theorem 4.1(ii) is independent of n). Thus, very large systems are not necessarily required for self-sustaining random autocatalysis, a concern that had been raised by Szathmáry (2000).

To establish Theorem 4.1 we require first two further results—Lemma 4.3 and Proposition 4.4, and to describe them we introduce a further definition.

We say that a reaction $r \in \mathscr{R}(n)$ is *globally catalysed* (or GC) if there exists any molecule in $X(n)$ that catalyses r. By the assumptions (R1) and (R2) above the probability that any forward reaction r is GC does not depend on r. Let p_* denote this probability and let $q_* = 1 - p_*$.

We will show that when p_* is sufficiently large then there exists a RAF $\mathscr{R} \subseteq \mathscr{R}_+(n)$ such that $X(n) - F \subseteq \pi(\mathscr{R})$—in other words all molecules that are not already supplied by F can be generated by catalysed reactions.

On the other hand, we will show that when p_* is small enough then the probability that there exists any globally catalysed reaction that generates any molecule from $X(t+1)$ from molecules in $X(t)$ is small—thus proving that the probability that a RAF exists is small.

The first step is to estimate the probability of global catalysation.

Lemma 4.3. *Consider the system $\mathscr{Q}(n)$ satisfying properties* (R1) *and* (R2) *and with $F = X(t)$ for a fixed t, and let $\lambda > 0$ be any positive constant.*

(i) *The probability q_* that a reaction $r \in \mathscr{R}_+(n)$ is not globally catalysed is given by*

$$q_* = \prod_{x\in X(n)} \left(1 - \frac{\mu_n(x)}{r_n}\right).$$

In particular,

(ii) *if $\mu_n(x) \leqslant \lambda n$ for all x then*

$$q_* \geqslant \exp\left(-\lambda\left(1 + O\left(\frac{1}{n}\right)\right)\right),$$

(iii) *if $\mu_n(x) \geqslant \lambda n$ for all x then*

$$q_* < \mathrm{e}^{-\lambda},$$

(iv) *if $\mu_n(x) \geqslant \lambda\theta_n|x|$ for all x (where θ_n is as defined in Theorem* 4.1) *then*

$$q_* < \mathrm{e}^{-\lambda}.$$

Proof. (i) is immediate from (R1) and (R2). (ii) follows by combining part (i) and (4) to give

$$q_* \geqslant \left(1 - \frac{\lambda n}{r_n}\right)^{x_n} \geqslant \left(1 - \frac{\lambda}{x_n\left(1 - O\left(\frac{1}{n}\right)\right)}\right)^{x_n} = \exp\left(-\lambda\left(1 + O\left(\frac{1}{n}\right)\right)\right).$$

(iii) follow from (i) together with (4) which gives

$$q_* \leqslant \left(1 - \frac{\lambda n}{r_n}\right)^{x_n} \leqslant \left(1 - \frac{\lambda}{x_n}\right)^{x_n} < \mathrm{e}^{-\lambda},$$

as required. For (iv), combine part (i), the identity $|\{x \in X(n) : |x| = s\}| = \kappa^s$, and the inequality $(1-a)^b \leqslant \exp(-ab)$ for $a, b > 0$, to obtain

$$q_* \leqslant \prod_{s=1}^{n} \left(1 - \frac{\lambda s}{r_n}\right)^{\kappa^s} \leqslant \prod_{s=1}^{n} \exp\left(-\frac{\lambda s\kappa^s}{r_n}\right) = \exp\left(-\frac{c}{r_n}\sum_{s=1}^{n} s\kappa^s\right).$$

Now, $\sum_{s=1}^{n} s\kappa^s = r_{n+1}/\kappa$ from (3), and (iv) now follows by identifying θ_n with $\frac{r_{n+1}}{r_n\kappa}$ (again using (3)). Note that θ_n converges to 1 as $n \to \infty$. □

Proposition 4.4. *Consider a random catalytic reaction system $\mathscr{Q}(n)$ satisfying properties* (R1) *and* (R2) *and with $F = X(t)$ for a fixed t, where $t < n$. As before, denote the probability that a forward reaction is not globally catalysed by q_*. Then*

(i) *The probability that $\mathscr{Q}(n)$ has a RAF is at most*

$$1 - q_*^{2x_t^2}$$

(ii) *If* $\kappa q_* < 1$ *then the probability that* $\mathcal{Q}(n)$ *has a RAF* $\mathcal{R}$ *with* $X(n) - F \subseteq \pi(\mathcal{R})$ *is at least*

$$1 - \frac{\kappa(\kappa q_*)^t}{1 - \kappa q_*}.$$

Proof. (i) Note that there are at most $2x_t^2$ forward reactions whose reactants (inputs) lie in $X(t)$. With probability at least $q_*^{2x_t^2}$ none of these reactions is GC, in which case there is no RAF for the system. The first part of the proposition now follows.

(ii) Note that, for any $s \geqslant t$ the probability that a molecule x with $|x| = s+1$ is not generated by any forward GC reaction from $X(s)$ is given by q_*^s. Therefore, the expected number of molecules x with $|x| = s+1$ which are not generated by a forward GC reaction is $\kappa^{s+1} q_*^s$. In particular, the probability that there is a molecule in $X(s+1)$ that is not generated by a forward reaction from $X(s)$ is at most $\kappa^{s+1} q_*^s$. This in turn implies that the probability that all molecules in $X(n)$ are generated by forward GC reactions is at least

$$1 - \kappa \sum_{s=t}^{n} (\kappa q_*)^s \geqslant 1 - \kappa \sum_{s=t}^{\infty} (\kappa q_*)^s = 1 - \frac{\kappa(\kappa q_*)^t}{1 - \kappa q_*}.$$

Finally, note that if all molecules in $X(n) - F$ are generated by a set $\mathcal{R}$ of forward GC reactions, and since $t < n$ (so that condition (iii) in the definition of an RAF is satisfied) we have that $\mathcal{R}$ is a RAF for $\mathcal{Q}(n)$. □

Proof of Theorem 4.1. (i) By Proposition 4.4 (i) the probability that $\mathcal{Q}(n)$ has a RAF is at most $1 - q_*^{2x_t^2}$ which by Lemma 4.3(ii) is at most

$$1 - \left[\exp\left(-\lambda\left(1 + O\left(\frac{1}{n}\right)\right)\right)\right]^{2x_t^2} = 1 - \exp\left(-2\lambda x_t^2\left(1 + O\left(\frac{1}{n}\right)\right)\right).$$

Clearly if $\mathcal{Q}(n)$ has no RAF, then it also has no Ω-complex RAF, for any $\Omega \subseteq 2^{X(n)-F}$.

(ii) This follows, by combining Proposition 4.4(ii) with Lemma 4.3 (iii) and (iv), and noting that a RAF $\mathcal{R}$ of $\mathcal{Q}(n)$ for which $X(n) - F \subseteq \pi(\mathcal{R})$ is also an Ω-complex RAF for any $\Omega \subseteq 2^{X(n)-F}$. □

5. An analogous result for CAFs

The degree of catalysation required for a CAF to arise in the system $\mathcal{Q}(n)$ is much greater than for a RAF. This seems reasonable since the definition of a CAF involves a much stronger requirement than a RAF on a set of reactions. However the extent of the difference is interesting, and is given by the following analogue of Theorem 4.1.

Theorem 5.1. *Consider the random catalytic reaction system* $\mathcal{Q}(n)$ *and suppose that* $F = X(t)$. *Let* $\lambda \geqslant 0$ *and let* $\Omega \subseteq 2^{X(n)-F}$.

(i) *If*

$$\mu_n(x) \leqslant \frac{\lambda}{x_t^3} \cdot r_n$$

for all $x \in X(n)$, *then the probability that* $\mathcal{Q}(n)$ *has a* Ω*-complex CAF is at most*

$$1 - \left(1 - \frac{\lambda}{x_t^3}\right)^{2x_t^3} \leqslant 2\lambda.$$

(ii) *If*

$$\mu_n(x) \geqslant \frac{\lambda}{x_t} \cdot r_n$$

for all $x \in X(n)$, *then the probability that* $\mathcal{Q}(n)$ *has a* Ω*-complex CAF is at least*

$$1 - \frac{\kappa(\kappa e^{-\lambda})^t}{1 - \kappa e^{-\lambda}}.$$

Before presenting the proof of this result, we note that while the degree of catalysation required for the likely occurrence of a RAF was that $\mu_n(x)$ should grow at least linearly with n (Theorem 4.1) the requirements for a CAF are quite different: by Theorem 5.1 $\mu_n(x)$ must grow at least linearly with r_n—and thereby *exponentially* with n.

Proof of Theorem 5.1. (i) Let $\mathcal{R}' := \{r \in \mathcal{R}_+(n) : \rho(r) \subseteq F\}$, the set of all forward reactions that have all their reactants in F. The probability that any given reaction $r \in \mathcal{R}'$ is not catalysed by at least one element of F is given by

$$\prod_{x \in F}\left(1 - \frac{\mu_n(x)}{r_n}\right).$$

Thus, the probability that none of the reactions in $\mathcal{R}'$ are catalysed by at least one element of F is

$$\left(\prod_{x \in F}\left(1 - \frac{\mu_n(x)}{r_n}\right)\right)^{|\mathcal{R}'|}.$$

In particular if $\mu_n(x) \leqslant \frac{\lambda r_n}{x_t^3}$, then, since $|F| = x_t$ and $|\mathcal{R}'| \leqslant 2x_t^2$, this probability (that none of the reactions in $\mathcal{R}'$ is catalysed by at least one element of F) is at least

$$\left(1 - \frac{\lambda}{x_t^3}\right)^{2x_t^3} \geqslant 1 - 2\lambda. \tag{5}$$

However when none of the reactions in $\mathcal{R}'$ is catalysed, then $\mathcal{Q}(n)$ does not have a CAF. Thus, the probability that $\mathcal{Q}(n)$ has a CAF is at most 1 minus the expression in (5), as required.

(ii) For every molecule in $x \in X(n)$, and each $s \in \{t, \ldots, n\}$ let $E_s(x)$ be the event that there is at least one reaction r_x of the form $a + b \to x$, where $a, b \in X(s)$, that is catalysed by at least one molecule in $X(s)$.

Now, if $\mu_n(x) \geqslant \frac{\lambda r_n}{x_t}$, then for any forward reaction r, the probability that r is not catalysed by at least one molecule in $X(s)$ (for $s \geqslant t$) is at most

$$\left(1 - \frac{\lambda}{x_t}\right)^{x_s} \leqslant \exp\left(-\lambda \frac{x_s}{x_t}\right) \leqslant e^{-\lambda}$$

and since, for each x there are $|x| - 1$ choices for r_x we have

$$\mathbb{P}(E_s(x)^c) \leqslant \exp(-\lambda(|x| - 1)), \tag{6}$$

where $E_s(x)^c$ is the complementary event to $E_s(x)$.

Consider the event

$$E_s := \bigcap_{x \in X(s+1) \setminus X(s)} E_s(x).$$

By (6) and the identity $|X(s+1) - X(s)| = \kappa^{s+1}$ we have $P(E_s^c) < \kappa^{s+1} e^{-\lambda s}$ and so

$$\mathbb{P}\left(\bigcap_{s=t}^{n-t} E_s\right) \geqslant 1 - \sum_{s=t}^{\infty} \kappa^{s+1} e^{-\lambda s} = 1 - \frac{\kappa(\kappa e^{-\lambda})^t}{1 - \kappa e^{-\lambda}}.$$

However the event $\bigcap_{s=t}^{n-t} E_s$ ensures that the nested collection of reactions $\mathscr{R}_i := \{r_x : x \in X(t+i)\}$, $i = 1, \ldots, n-t$ forms a CAF for $\mathscr{Q}(n)$, and moreover, one for which the maximal set $\mathscr{R}_{n-t}$ generates all elements of $X(n) - F$—thus it is also a Ω-complex CAF for any $\Omega \subseteq 2^{X(n)-F}$. This completes the proof. □

6. Discussion

The question of how life first arose on earth is a multifaceted problem that stands out as one of the major questions in science (see for example Dyson, 1985; Fenchel, 2002; Joyce, 1989; Szathmáry, 1999; Szathmáry and Maynard Smith, 1997). One dilemma, frequently dubbed the 'chicken and egg' problem is the question of which (if either) came first: hereditary (molecules such as DNA or RNA that carry information but do not easily catalyse reactions), or metabolism (proteins that carry out reactions but do not replicate). An alternative possibility is than an autocatalytic system of molecules including RNA and proteins and possibly other molecules emerged as the first primitive prebiotic system. The theoretical investigation of catalytic reaction systems is an attempt to address just one aspect of this theory. This concerns the issue of whether, as Kauffman has maintained, we should expect self-sustaining, autocatalytic networks to emerge in random chemical systems once some threshold (in 'complexity', 'connectivity' or 'catalysation rate') is exceeded, or whether there is the requirement of some fine-tuning of the underlying biochemistry for such networks to occur. Orgel (1992) raises this as concern about autocatalytic network models commenting that "it is always difficult in such theoretical models to see how to close the cycle without making unreasonable assumptions about the specificity of catalysis."

Our results here have helped delineate precisely how much catalysation is required in order for random sequence-based chemical reaction systems (without any 'fine-tuning') to likely give rise to a RAF. In contrast to a CAF, where a high degree of catalysation is required when the maximal sequence length n is large, the likely occurrence of a RAF depends just on whether the catalysation function $\mu_n(x)$ grows sublinearly or super-linearly with n (Corollary 4.2). The techniques developed in this paper may provide some analytical predictive tools for biochemists design *in vitro* prebiotic experiments with large number of variants of RNA sequences and other molecules.

The development of a self-sustaining autocatalytic system would clearly be only one step towards life, in particular a reproducing system that is capable of undergoing Darwinian selection eventually needs to develop. Here, the recent concept of a 'Eigen–Darwin' cycle (Poole et al., 1998) may hold promise.

Questions for future work would be to explore how the results in this paper would be influenced by allowing random inhibitory catalysations, or side reactions that could destroy some of the crucial reactants (this problem has been referred to by Szathmáry (2000) as the "plague of side reactions"). This second phenomena can be formally regarded as a special case of the first, since if x is a reactant for a reaction r and x is degraded in the presence of another molecule y then we can (formally) regard y as inhibiting the reaction r. The model studied in this paper could also be refined to better suit the graph-theoretic properties of real metabolic networks which have recently been investigated (Jeong et al., 2000; Wagner and Fell, 2001).

Acknowledgements

The first author is supported by a Miller fellowship. The second author thanks the NZIMA (Maclauren Fellowship) for supporting this work. We also thank the referees for helpful comments on an earlier version of this manuscript.

Appendix A. Proof of Proposition 3.1

The decision problem is clearly in the class NP. To show it is NP-complete we provide a reduction from 3-SAT. Consider an expression P in conjunctive normal form involving binary variables $x_1, \ldots, x_n$ and where each clause in P involves at most three variables.

Thus we can write

$$P = C_1 \wedge C_2 \wedge \cdots \wedge C_k,$$

where

$$C_i = \bigvee_{j \in T(i)} x_j \bigvee_{j \in F(i)} \overline{x}_j$$

and $T(i), F(i) \subseteq \{1, \ldots, n\}$, $|T(i)| + |F(i)| = 3$.

Given P construct a catalytic reaction system $\mathscr{Q} = (X, \mathscr{R}, \gamma_{C(+),C(-)})$ as follows: let $F := \{x_1, \ldots, x_n\}$, let

$$X := \{x_1, \ldots, x_n, f_1, \ldots, f_n, t_1, \ldots, t_n, \theta_1, \ldots, \theta_k, 1\}.$$

Informally, x_i will correspond to the variable x_i in the formula; a reaction producing t_i (respectively, f_i) will be catalysed if the truth assignment of x_i is true (respectively, false), and the reaction producing θ_i will be catalysed if the ith clause is satisfied.

More formally we let $\mathscr{R} = \mathscr{R}_1 \cup \mathscr{R}_2 \cup \mathscr{R}_3$ where $\mathscr{R}_1$ consists of all reactions

$$x_i \xrightarrow{1(+),t_i(-)} f_i, \quad x_i \xrightarrow{1(+),f_i(-)} t_i$$

for $1 \leqslant i \leqslant n$. In words x_i is the sole reactant for f_i and t_i but f_i inhibits the catalysation of t_i and vice versa.

$\mathscr{R}_2$ consists of all reactions

$$\begin{cases} t_j \xrightarrow{1(+)} \theta_i & \text{if } j \in T(i), \\ f_j \xrightarrow{1(+)} \theta_i & \text{if } j \in F(i) \end{cases}$$

for $1 \leqslant j \leqslant n$ and $1 \leqslant i \leqslant k$. Finally $\mathscr{R}_3$ consists of the single reaction

$$\{\theta_1, \ldots, \theta_k\} \xrightarrow{1(+)} 1.$$

Now, we claim that P has a satisfying truth assignment if and only if $\mathscr{Q}$ has a RAF. To establish this, first assume that P has a satisfying assignment. Fix such an assignment z and let $\{T, F\}$ be a partition of $\{1, \ldots, n\}$ corresponding to the variables that are true (respectively, false) in z.

Now, consider $\mathscr{R}'_1 \cup \mathscr{R}'_2 \cup \mathscr{R}_3$ where $\mathscr{R}'_1 \subseteq \mathscr{R}_1$ consists of the reactions $x_i \to t_i$ for all $i \in T$, and the reactions $x_i \to f_i$ for all $i \in F$. $\mathscr{R}'_2$ will consist of the reactions $t_i \to \theta_j$ for all $i \in T \cap T(j)$ and $f_i \to \theta_j$ for all $i \in F \cap F(j)$. Since the assignment z satisfies the formula it follows that $\mathscr{R}'_1 \cup \mathscr{R}'_2 \cup \mathscr{R}_3$ is a RAF.

Next, we have to show that if the system has a RAF the formula has a satisfying truth assignment. Suppose the system has a RAF $\mathscr{R}'$. Clearly $\mathscr{R}_3 \subset \mathscr{R}'$. This in turn implies that the reactions producing $\theta_1, \ldots, \theta_k$ are all catalysed. Thus for all $1 \leqslant i \leqslant k$, there either exists some $j \in T(i)$ such that the reaction producing t_j is catalysed or there exists some $j \in F(i)$ such that the reaction producing f_j is catalyzed. Moreover, for all i at most one of the reactions producing t_i and f_i can be catalyzed. We now define z_i to be true if the reaction producing t_i is catalysed and false if the reaction producing f_i is catalyzed (z_i is defined arbitrarily otherwise). Then z is a satisfying assignment as required. □

References

Bollobas, B., 2001. Random Graphs, second ed. Cambridge Studies in Advanced Mathematics, vol. 73. Cambridge University Press, Cambridge.

Bollobas, B., Rasmussen, S., 1989. First cycles in random directed graph processes. Discrete Math. 75, 55–68.

Cohen, J.E., 1988. Threshold phenomena in random structures. Discrete Appl. Math. 19, 113–128.

Dyson, F., 1982. A model for the origin of life. J. Mol. Evol. 18, 344–350.

Dyson, F., 1985. Origins of Life. Cambridge University Press, Cambridge.

Eigen, M., 1971. Self-organisation of mater and evolution of biological macromolecules. Naturwissenshaften 58, 465–523.

Eigen, M., Schuster, P., 1979. The Hypercycle—A Principle of Natural Selforganisation. Springer, Heidelberg.

Erdös, P., Rényi, A., 1960. On the evolution of random graphs. Magyar Tud. Akad. Mat. Kut. Int. Közl 5, 17–61.

Fenchel, T., 2002. The Origin and Early Evolution of Life. Oxford University Press, Oxford.

Hordijk, W., Fontanari, J.F., 2003. Catalytic reaction sets, decay, and the preservation of information. Proceedings of KIMAS'03 (IEEE International Conference on Integration of Knowledge Intensive Multi-Agent Systems, Boston, MA), pp. 133–138.

Hordijk, W., Steel, M., 2004. Detecting autocatalyctic, self-sustaining sets in chemical reaction systems. J. Theor. Biol. 227 (4), 451–461.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, 2000. The large-scale organisation of metabolic networks. Nature 407, 651–654.

Joyce, G.F., 1989. RNA evolution and the origin of life. Nature 338, 217–224.

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

Kauffman, S.A., 1993. The Origins of Order: Self-organisation and Selection in Evolution. Oxford University Press, Oxford.

Lifson, S., 1997. On the crucial stages in the origin of animate matter. J. Mol. Evol. 44, 1–8.

Lohn, J.D., Comombano, S.P., Scargle, J., Stassinopoulos, D., Haith, G.L., 1998. Evolving catalytic reaction sets using genetic algorithms. Proceedings of the 1998 IEEE International Conference on Evolutionary Computation, Anchorage, AK, USA, pp. 487–492.

Maynard Smith, J., 1983. Models of evolution. Proc. Roy. Soc. London B 219, 315–325.

Maynard Smith, J., Szathmáry, E., 1995. The Major Transitions in Evolution. Oxford University Press, Oxford.

Orgel, L.E., 1992. Molecular Replication. Nature 358, 203–209.

Penny, D., 2004. An interpretive review of the origin of life research. Biol. Philos., in press.

Poole, A.M., Jeffares, D.C., Penny, D., 1998. The path from the RNA world. J. Mol. Evol. 46, 1–17.

Sharov, A.A., 1991. Self-reproducing systems: structure, niche relations and evolution. Biosystems 25, 237–249.

Scheuring, I., 2000. Avoiding catch-22 by early evolution of stepwise increase in copying fidelity. Selection 1, 135–145.

Steel, M., 2000. The emergence of a self-catalysing structure in abstract origin-of-life models. Appl. Math. Lett. 3, 91–95.

Szathmáry, E., 1999. The first replicators. In: Keller, L. (Ed.), Levels of Selection in Evolution. Princeton University Press, Princeton, NJ, pp. 31–52.

Szathmáry, E., 2000. The evolution of replicators. Philos. Trans. Roy. Soc. London B 355, 1669–1676.

Szathmáry, E., Maynard Smith, J., 1997. From replicators to reproducers: the first major transitions leading to life. J. Theor. Biol. 187, 555–571.

Wagner, A., Fell, D.A., 2001. The small world inside large metabolic networks. Proc. Roy. Soc. London B 268, 1803–1810.

Wills, P., Henderson, L., 1997. Self-organisation and information-carrying capacity of collectively autocatalytic sets of polymers: ligations systems. In: Bar-Yam, Y. (Ed.) Proceedings of the International Conference on Complex Systems, New England Complex Systems Institute.

Zintzaras, E., Santos, M., Szathmáry, E., 2002. Living under the challenge of information decay: the stochastic corrector model vs hypercycles. J. Theor. Biol. 217, 167–181.