Orig Life Evol Biosph (2006) 36:381–389
DOI 10.1007/s11084-006-9014-9

Emergent robustness in competition between autocatalytic chemical networks

Richard A. Goldstein

Received: 18 December 2005 / Accepted: 3 March 2006

Abstract The origin of auto-catalytic networks has been proposed as an initial step in pre-biotic evolution. It is possible to derive simple models where auto-catalytic networks naturally arise from simple chemical mixtures. In order for such a system to develop, there needs to be some degree of stability, what is characterised as 'robustness'. We demonstrate that competing systems generate this robustness as they create a distributed network of catalytic pathways.

Keywords Biogenesis · Chemical models · Molecular evolution · Origin of life · Population dynamics · Pre-biotic evolution

Introduction

Theories about the origin of life can generally be classified into the 'replication first' and 'metabolism first' approaches. The 'replication first' approach imagines the first proto-life forms were polymers capable of self-replication, generally RNA (Gilbert, 1986; Sievers and von-Kiedrowski, 1994; Bolli *et al.*, 1997). RNA's potential for catalysis (Kruger *et al.*, 1982) has suggested that metabolism could evolve in such an 'RNA world'. A major problem with this approach involves the Eigen paradox (Eigen, 1971; Maynard-Smith and Szathmáry, 1995; Scheuring, 2000). Errors in replication limit the size of a polymer that can be replicated without the loss of genetic information. Proto-nucleic acids could only be 10–100 residues long, too small to encode the replicases necessary for replication. One proposed solution is the development of 'hypercycles' of RNA – interactions of a number of different short self-replicating RNA molecules, each catalysing the replication of another.

An alternative is the 'metabolism first' approach, where self-sustaining chemical reactions occur before the origin of replication (Oparin, 1953; Oparin and Gladilin, 1980; Kauffman, 1986; Fontana and Buss, 1994; Dyson, 1999; Segre *et al.*, 2000; Segre and Lancet, 2000). By the rise of auto-catalytic networks (ACNs), these chemical reactions could develop, expand,

R. A. Goldstein (✉)
Mathematical Biology, National Institute of Medical Research, Mill Hill, NW7 1AA, UK
e-mail: richard.goldstein@nimr.mrc.ac.uk

and 'replicate' themselves in the sense of producing more systems containing these same autocatalytic networks. One specific model is the 'garbage bag model', which imagines that there are 'bags' formed by, as an example, amphipathic molecules, containing or composed of ACNs but permeable to the reactants necessary for the autocatalytic networks to continue in an out-of-equilibrium state. If the ACNs result in an expansion of the contents of the bag, the bag could divide in a manner similar to mitosis. In this case, the 'information' passed from parent to daughter would consist of the 'composome', that is, the relative concentrations of the various members of the ACNs (Segre *et al.*, 2000; Segre and Lancet, 2000). A 'mutation' in the composome, such as the addition of a new chemical species into the ACNs, would be inherited by the daughters. RNA might represent the 'first parasite', taking advantage of the pre-existing pre-evolved ACN for its own replication.

It is not difficult to describe a how an ACN can develop. For instance, Jain and Krishna (Jain and Krishna, 2001; Jain and Krishna, 2002) developed a specific model where different chemical species either work for the creation or annihilation of the other chemical species. The 'Jain-Krishna' dynamics proceeds on two different time scales. The rapid time-scale involves the generation of a steady-state equilibrium in the concentration of the various chemical species. The slower process is the changing in the parameters in the model, allowing different reactants to be removed or added to the system. They demonstrated that ACNs arise naturally in these dynamics, gradually extending to encompass essentially all of the various chemical species. Generally, the 'core' of the ACN is small – a few chemical species which form a small autocatalytic network - surrounded by other 'parasitic' species whose formation is catalysed by either one of these core species or another parasite but do not contribute to the catalysis of any of the core members of the ACN, either directly or indirectly. Unfortunately, these networks are fragile. After a period of time, one of the essential ('keystone') core members of the ACN is lost, the autocatalytic cycles are broken, and the network collapses.

It is difficult to see how transient ACNs can be the initial starting material for the development of life. Given this fragility, how is it possible to develop robust ACNs, that is, ACNs that do not so quickly lose an essential core member and collapse as a response to the removal and addition of various reactants? If such a development is possible, how does this robustness arise? What is its nature?

The 'replication first' hypercycle models are also susceptible to collapse. There can be the rise of cheaters, parasites, and short-cuts. One solution to this problem is by having selection occur at a larger level, by compartmentalising hypercycles in membrane vesicles and letting selection occur between vesicles (Eigen *et al.*, 1981; Maynard-Smith and Szathmáry, 1995). In this way, hypercycles infected with parasites would be eliminated. 'Altruistic' members of hypercycles can be selected through population dynamics similar to 'kinship selection'.

The same mechanism of selective pressure occurring at a higher level might be able to generate robust ACNs. In this paper, we describe what happens to a chemical ACN model when we have selection occurring at the garbage-bag level. We observe a rapid rise in the robustness of the ACNs, corresponding to an increase in catalytic formation. In particular, we find a rapid increase in the complexity of the ACNs, including a decrease in the number of parasitic chemical species. The ACNs develop distributed robustness, where the robustness results from the overall complexity of the catalytic network.

Methods

We use the Jain-Krishna chemical system evolution model (Jain and Krishna, 2001; Jain and Krishna, 2002). In this model, there are $N = 100$ different chemical species that

can either catalyse the formation or destruction of any of the other species, or of themselves, with the exception that no chemical species is allowed to be autocatalytic, that is, to catalyse its own formation. The relative concentration of the N species is given by $\{x_1, x_2, x_3 ... x_N\}$ ($\sum_i x_i = 1,\ x_i \geq 0 \,|\, \forall i$). These concentrations change according to the formula

$$x_i(t + \Delta t) = (1 - \lambda) \max \left(x_i(t) + \sum_j c_{ij} x_j, 0 \right), \tag{1}$$

where λ represents the dilution to maintain $\sum_i x_i = 1$, and thus represents the increase of the system. A positive value of c_{ij} means that species j catalyses the formation of species i, while a negative value means j catalyses the destruction of i. (An alternative interpretation is that these represent how much chemical species i catalyses the process of species j joining or leaving of system (Segre *et al.*, 2000).) The c_{ij} values are chosen at random; only a small fraction ($P = 0.005$) of c_{ij} values are chosen to be non-zero, with the non-zero values chosen randomly from a uniform distribution in the interval $[-1, 1]$, except for non-zero c_{ii} values which are chosen randomly from a uniform distribution in the interval $[-1, 0]$ (i.e. no autocatalytic chemical species). There are presumably a number of other simple molecules that serve as reactants in the various catalytic reactions. It is assumed that these are fed into the network through a steady diffusive flux.

As discussed in the Introduction, the model involves chemical dynamics occurring on a fast time-scale, combined with evolutionary dynamics occurring on a much slower time-scale. For a given set of $\{c_{ij}\}$, the chemical dynamics represented by Equation (1) are continued until convergence. This then represents the 'steady-state' of the system as described by $\{c_{ij}\}$. For the evolutionary dynamics, the chemical species with the smallest value of x_k is eliminated, and replaced with a new chemical species with a small (0.0001) initial concentration. The various coefficients representing the properties of this new species (c_{ik} and c_{kj}) are again chosen randomly. The chemical dynamics are then repeated until convergence. Each cycle of replacement of the rarest chemical species and establishment of the new steady-state represents one generation.

Two different types of evolutionary runs were simulated. The first involved the simple dynamics described above, run for 100,000 generations (independent evolution). A total of 2000 such simulations were performed. The second type involved a *population* of 100 chemical systems undergoing Jain-Krishna dynamics, with these systems competing against each other (competing systems). An initial system was created as described above and replicated 100 times, forming the initial generation. In each system the chemical species with the smallest concentration was eliminated and replaced with a new chemical species with random values of c_{ik} and c_{kj}, and chemical dynamics were simulated until convergence. In order to represent the stochastic nature of the system population dynamics, the next generation of systems were chosen at random from the systems in the current generation, with replacement, with relative probability equal to the fitness of each system. The 'fitness' of each system was defined as the final steady-state value of λ, that is, the excess chemical species that had to be drained away in order to maintain a constant total chemical concentration, equal to the rate of excess production of the ACS. This systems population dynamics continued for 100,000 generations. 20 such simulations were performed.

In the competing systems simulations there are three different dynamics processes: chemical dynamics on an extremely rapid time-scale, and evolution of the different chemical systems as well as system population dynamics occurring on a slower time-scale.

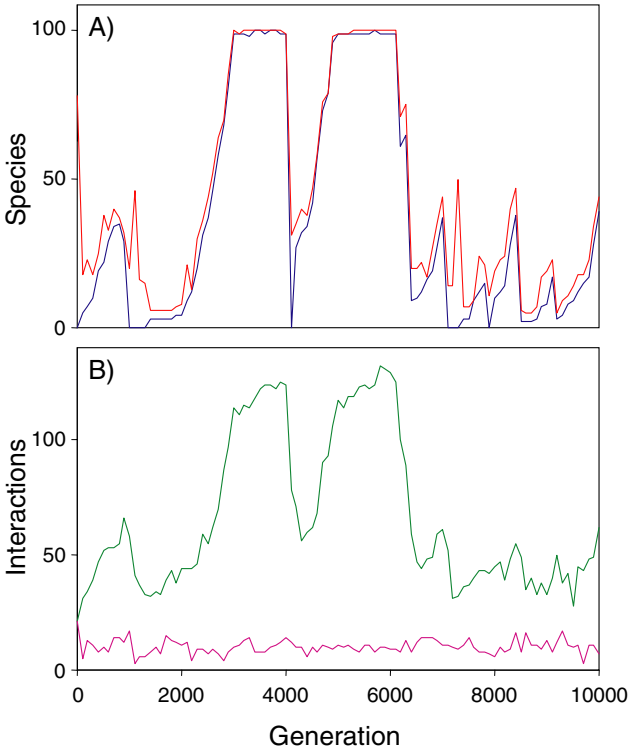

Fig. 1 Dynamics of a discrete chemical systems undergoing Jain-Krishna dynamics (Jain and Krishna, 2001; Jain and Krishna, 2002). (A) The total number of chemical species with non-negligible concentrations (red) and the total number involved in the ACS (blue). (B) Number of positive (green) and negative (magenta) catalytic interactions in the system

Results

A typical simulation for a chemical system undergoing independent Jain-Krishna dynamics is shown in Figure 1. As described, there is a random formation of an ACS, involving a small number of chemical species. Once this ACS is formed, other chemical species add to the ACS, generally in a parasitic manner. This continues until the vast majority of the species are members of the ACS. The broad participation of nearly all of the chemical species, with non-negligible concentrations, was a characteristic feature of the presence of an ACS. There is a correspondingly large increase in the number of positive interactions in the system, as species whose formation is not catalysed by a member of the ACS are generally eliminated. The number of negative interactions remains roughly constant. The ACS continues until one of the key members of the ACS 'core' is eliminated, and the ACS collapses. Some time later, a similar or dissimilar ACS may form, again lasting for a period of time before collapsing.

After an initial period of development of approximately 5000 generations, the independent evolution simulations had a roughly constant fraction (18%) of systems in the ACS state. Conversely, with the competing systems simulations, after approximately 3000 generations essentially all systems were in the ACS state. ACS states in general have higher fitnesses (λ) than non-ACS states, so any system with a collapsed ACS would be rapidly removed from the population of systems.

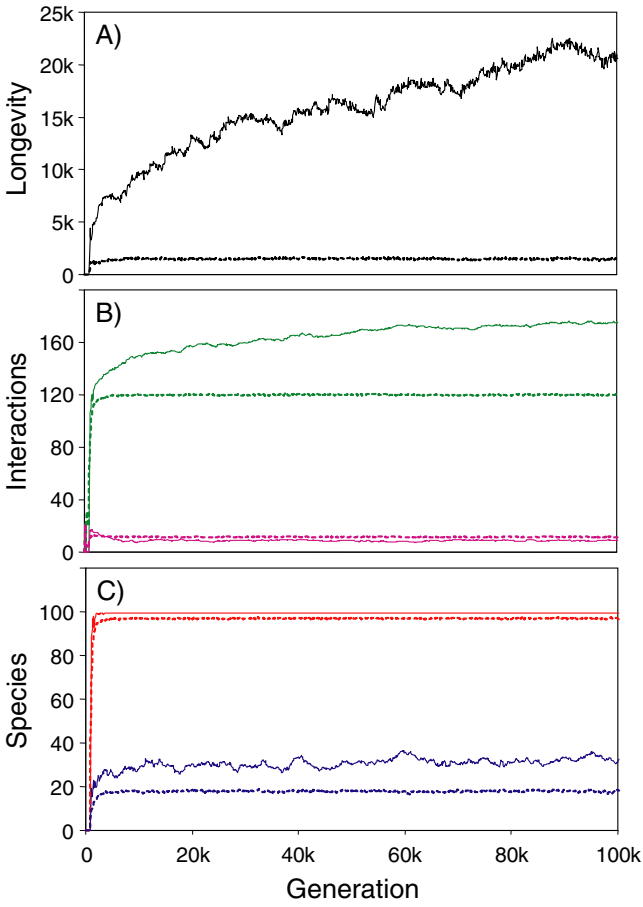

Fig. 2 Comparison of the average values of various quantities resulting from competing populations of chemical systems (solid) and independent evolution (dashed). Only simulations with ACSs were included in the averages. (A) Average longevity, that is, the mean length of time an active ACS can undergo independent evolution before the collapse of the ACS. (B) Average number of positive (green) and negative (magenta) catalytic interactions in the system. (C) Average number of participating members of the ACS (red), compared with the number of non-parasitic members of the ACS (blue). Parasitic members were defined recursively as those that do not catalyse the formation of any chemical species except for those previously designated as parasitic

We can calculate the 'longevity' of a given chemical system by creating multiple copies of the system, and having each undergo independent chemical dynamics until the ACS collapsed. (The ACS was considered to be still present as long as 80% of the chemical species had non-negligible concentrations.) The longevity is defined as the average time to collapse of that system. As shown in Figure 2, the average longevity of the independent evolution systems rapidly increases to approximately 1500 generations, where it remained for the rest of the simulation. Conversely, the average longevity of the systems resulting from the competing systems simulations maintains a strong upward trend, reaching over 20,000 generations by the end of the simulation. This corresponded to a large increase in the average number of positive interactions in resulting from the competing systems simulations, increasing to around 175,

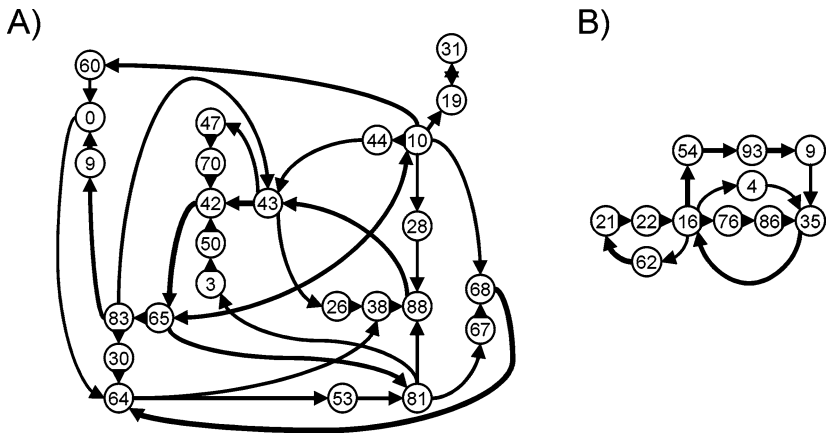

Fig. 3 Two ACSs resulting from competitive population dynamics between chemical systems (A) and from independent evolution (B), showing the positive (catalysing) reactions, with relative strength (magnitude of c_{ij}) represented by the width of the connecting vector. Both ACSs had the same total number of positive catalytic interactions. Parasitic species have been removed for clarity

in comparison with the average number of positive interactions in the independent evolution systems which plateaued at approximately 120. The average numbers of negative interactions in these two groups were similar.

As shown in Figure 2, the average number of species involved in the ACS states is not dissimilar in the two different types of simulations, being slightly higher for the systems undergoing inter-system competition (99.2) than for systems evolving independently (96.9). The majority of involved species in both cases were parasitic, defined recursively as those that do not catalyse the formation of any chemical species except for those previously designated as parasitic. There was, however, a significant difference – on average approximately 32 of the species in systems involved in competition were non-parasitic, compared with only 18 of the chemical species in the individually-evolving systems. It was this increase in the non-parasitic 'core' of the ACS that might provide much of the robustness.

Figure 3 shows two ACS cycles, one from the set of competing systems, the other from the independently-evolving systems, from generation 10,000. These two were chosen so that the two ACSs had the same number of positive interactions (132), yet the longevity of the ACS resulting from inter-system competition was significantly longer (4250) than the longevity of the ACS from the independently evolving system (980). In Figure 3 the parasitic residues have been removed. It is clear how much larger and more extensive the competing systems ACS is, with 25 non-parasitic species in the results of the population evolution compared with the 11 in the independent evolution system ACS. Each non-parasitic species in the competing systems ACS catalyses (and is catalysed by) on average 1.5 other non-parasitic species, compared with 1.3 for the independent evolution system. In fact, there is a single species in the independent evolution ACS (16) that, if removed, would destroy the ACS.

Figure 4 shows how the concentration of the various species evolves during two typical runs by measuring the similarity of the concentration vectors at different times in the simulation. The periods of ACS in the independent evolution simulations appear as small 'blocks' along the diagonal. It is clear that different ACSs occurring during the independent evolution simulation have relatively little similarity. In contrast, there is a clear longevity in the ACSs in the competing systems simulations. Figure 5 shows the strong correlation between fitness and longevity, both for the independent evolution and competing systems runs.

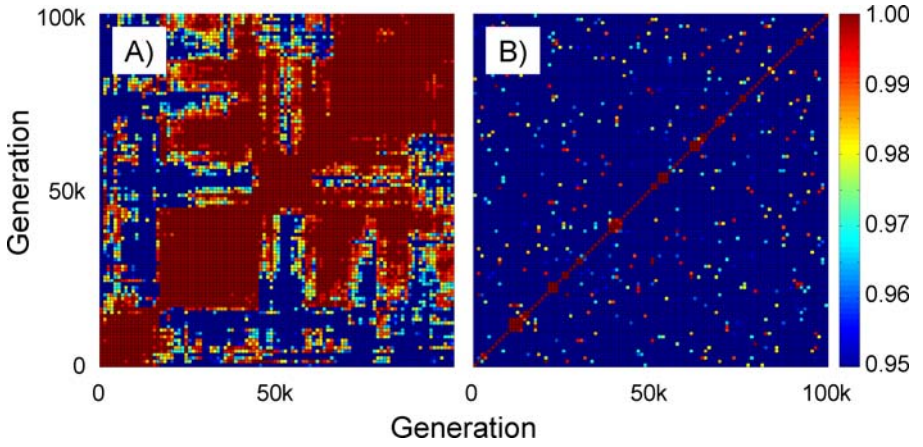

Fig. 4 Representation of changes in the chemical species concentrations, measured as similarities between the concentrations between different time-points in the simulations. (A) Evolution of a member of a set of chemical systems undergoing population dynamics. (B) Evolution of a chemical system undergoing independent evolution. The simulation represented in (B) is the same as that represented in Figure 1. (Note the difference in scales.) Similarity between two concentration vectors $\vec{x_i}$ and $\vec{x_j}$ was measured by first computing the scaler-product of the concentrations $\vec{x_i} \cdot \vec{x_j}$ (Segre *et al.*, 2000). The entries of one of the vectors were then randomly permuted 1000 times, and the scaler-product computed for each permutation. Figure 4 shows the fraction of the permutations that have smaller dot-products than that of the original pair of concentrations

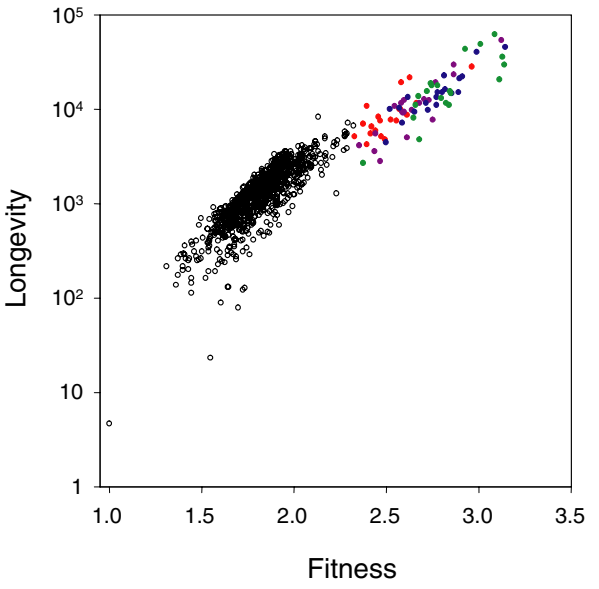

Fig. 5 Relationship between fitness (λ) and longevity for systems with ACSs. Independently-evolving systems at generation 100,000 (open circles) are compared with systems undergoing competitive population dynamics at generation 10,000 (red), 20,000 (purple), 50,000 (blue), and 100,000 (green)

Discussion

In the individual evolution of discrete chemical systems, we observe, as did Jain and Krishna (Jain and Krishna, 2001; Jain and Krishna, 2002), a rise and collapse of autocatalytic networks. As the ACN expand to include essentially all chemical species, the core members of the network do not have any specific advantage over their parasitic neighbours. It is only a

matter of time before one of these core members are removed by the evolutionary dynamics and the network collapses.

Conversely, when different systems compete with each other in a population context, there is a rapid rise in robust ACN systems. There is an increase in the number of positive (catalysing) interactions. There is also an increase in the complexity of the ACNs, measured by the number of core members of the network. As a result, the ACNs are able to survive for much longer, including in subsequent evolution as a discrete chemical system. The ACNs have become robust. As demonstrated in Figure 4, the ACS also has a strong tendency to remain in a relatively consistent state, lasting on the order of 30,000 generations. In comparison, the transient ACSs formed by chemical systems in the independent evolution simulations have statistically-insignificant similarity to each other.

How does this robustness arise? There are generally three possible mechanisms for its origin. Firstly, it might be that robustness is its own reward, that there may be selection pressure acting at the individual level for robustness. As an example, a certain level of robustness can result in an organism that is better able to survive changes in the external environment or normal statistical fluctuations in the internal state. Secondly, robustness might be correlated with some other important property. In this manner, robustness might be a fortuitous 'hitchhiking' quality. The third is a consequence of Eigen's quasispecies model (Eigen and McCaskill, 1988; Eigen *et al.*, 1989); the fact that fitness is calculated over a polymorphic ensemble of different species results in an increased fitness for flatter regions of the fitness landscape (Schuster and Swetina, 1988; Bornberg-Bauer and Chan, 1999; Nimwegen *et al.*, 1999; Wilke, 2001a,b; Taverna and Goldstein, 2002).

The mechanism of the robustness in this situation seems to lie with the second explanation. As is demonstrated in Figure 5, longevity is highly correlated with fitness, even among systems evolving independently where there is no selective pressure either for longevity or for fitness. The selective pressure for increased fitness in the systems undergoing competitive population dynamics result in more complex ACNs, with more positive interactions. In this manner, there is selection for complexity, which is correlated with the greater degree of robustness, as measured by the increased longevity.

Robustness can generally arise either through direct redundancy, that is, duplication of essential elements, or through 'distributed robustness' where the robustness is the result of the network topology and architecture; the robustness of biological systems often comes from this distributed robustness (Wagner, 2005). In the population dynamics of competing chemical systems, as is clear from Figure 5, involves distributed robustness – no chemical species share the same catalytic reactions. It is rather the dense network of catalytic reactions that results in a distribution of importance among a number of different species.

References

Bolli M, Micura R, Eschenmoser A (1997) Pyranosyl-Rna: chiroselective self-assembly of base sequences by ligative oligomerization of tetranucleotide-2′,3′-cyclophosphates (with a Commentary Concerning the Origin of Biomolecular Homochirality). Chem Biol 4:309–320

Bornberg-Bauer E, Chan HS (1999) Modeling evolutionary landscapes: mutational stability, topology, and superfunnels in sequence space. Proc Nat'l Acad Sci USA 96:10689–10694

Dyson F (1999) Origins of Life Cambridge University Press, Cambridge

Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 10:465–523

Eigen M, McCaskill J (1988) Molecular quasi-species. J Phys Chem 92:6881–6891

Eigen M, McCaskill J. Schuster P (1989) The molecular quasi-species. Adv Chem Phys 75:149–263

Eigen M, Schuster P, Gardiner W, Winkler-Oswatitsch R (1981) The origin of genetic information. American Scientist 244:78–94
Fontana W, Buss LW (1994) What would be conserved if "the tape were played twice"?. Proc Natl Acad Sci USA 91:757–761
Gilbert W (1986) The RNA world. Nature 319:618
Jain S, Krishna S (2001) A model for the emergence of cooperation, interdependence, and structure in evolving networks. Proc Nat'l Acad Sci USA 98:543–547
Jain S, Krishna S (2002), Large extinctions in an evolutionary model: the role of innovation and keystone species. Proc Nat'l Acad Sci USA 99:2055–2060
Kauffman SA (1986) Autocatalytic sets of proteins. J Theor Biol 119:1–24
Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. Cell 31:147–157
Maynard-Smith J, Szathmáry E (1995) The major transitions of evolution. Freeman, Oxford.
Nimwegen Ev, Crutchfield JP, Huynes M (1999) Neutral evolution of mutational robustness. Proc Nat'l Acad Sci USA 96:9716–9720
Oparin AI (1953) The origin of Lilfe. Dover, New York
Oparin AI, Gladilin KL (1980) Evolution of self-assembly of probionts. Biosystems 12:133–145
Scheuring I (2000) Avoiding atch-22 of early evolution by stepwise increase in copying fidelity. Selection 1–3:135–145
Schuster P, Swetina J (1988) Stationary mutant distributions and evolutionary optimization. Bull Math Biol 50:635–660
Segre D, Ben-Eli D, Lancet D (2000) Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. Proc Nat'l Acad Sci USA 97:4112–4117
Segre D, Lancet D (2000) Composing life. EMBO Reports 1:217–222
Sievers D, von-Kiedrowski G (1994) Self-replication of complementary nucleotide-based oligomers. Nature 369:221–224
Taverna DM, Goldstein RA (2002) Why are proteins so robust to site mutations. Journal of Molecular Biology 315:479–484
Wagner A (2005) Distributed robustness versus redundancy as causes of mutational robustness. Bioessays 27:176–188
Wilke CO (2001a) Adaptive evolution on neutral networks. Bull Math Biol 63:715–730
Wilke CO (2001b) Selection for fitness vs. selection for robustness in rna secondary structure folding. Evolution 55:2412–2420