Molecular BioSystems

REVIEW

View Article Online
View Journal

Cite this: DOI: 10.1039/c5mb00593k

Prebiotic network evolution: six key parameters

Philippe Nghe,†[a] Wim Hordijk,†[b] Stuart A. Kauffman,[c] Sara I. Walker,[d] Francis J. Schmidt,[e] Harry Kemble,[a] Jessica A. M. Yeates[f] and Niles Lehman*[f]

Received 3rd September 2015,
Accepted 12th October 2015

DOI: 10.1039/c5mb00593k

www.rsc.org/molecularbiosystems

The origins of life likely required the cooperation among a set of molecular species interacting in a network. If so, then the earliest modes of evolutionary change would have been governed by the manners and mechanisms by which networks change their compositions over time. For molecular events, especially those in a pre-biological setting, these mechanisms have rarely been considered. We are only recently learning to apply the results of mathematical analyses of network dynamics to prebiotic events. Here, we attempt to forge connections between such analyses and the current state of knowledge in prebiotic chemistry. Of the many possible influences that could direct primordial network, six parameters emerge as the most influential when one considers the molecular characteristics of the best candidates for the emergence of biological information: polypeptides, RNA-like polymers, and lipids. These parameters are viable cores, connectivity kinetics, information control, scalability, resource availability, and compartmentalization. These parameters, both individually and jointly, guide the aggregate evolution of collectively autocatalytic sets. We are now in a position to translate these conclusions into a laboratory setting and test empirically the dynamics of prebiotic network evolution.

Abiogenesis is an unresolved phenomenon that has received significant attention.[1–10] In the past half century the focus of much research has been on the elucidation of mechanisms by which self-replicating molecules could have arisen from some type of prebiotic environment.[11–14] In recent years however, excitement has been generated by a somewhat alternative viewpoint, that a network of interdependent molecular species could have sparked the transition from chemistry to biology. In this view, the critical unit of origin and evolution is a cooperative collection of molecules rather than a single "selfish" replicating entity. Such a collection has been proposed in a variety of frameworks. For example, Gánti in one of the earliest descriptions of the protocell, proposed a chemoton in which the molecular components responsible for replication and metabolism together formed an autocatalytic system.[15] Eigen and Schuster envisioned a hypercycle in which cooperative interactions among its components led to a dramatic (hyperbolic) growth of each member.[16–19] In parallel Kauffman put forth the notion of a collective autocatalytic set (CAS),[20,21] a system whose critical characteristic is that catalytic closure of all components could be possible such that the set could self-replicate. The CAS concept has been extended and formalized as a reflexively autocatalytic and food-generated (RAF) set by Hordijk and Steel.[22] There, the importance of the environment in providing raw materials (*i.e.*, food) for the network was made mathematically explicit. These constructs typically refer to information-bearing networks, rather than the energy-producing ones such as those discussed by Wächtershäuser, Russell, and others.[23–25] But they should also apply to metabolic cycles – even if networks may be easier to conceptualize in the viewpoint that the critical transition for life was the advent of information-rich biopolymers;[16,26,27] similar organizational principles may apply in both cases.

The idea of a network as an early manifestation of life brings several advantages. First, as noted earlier,[20,21,28] networks allow distributed function. Not all catalytic function need reside in any one component, so long as the network (CAS) as a whole contains enough functionality to allow for the synthesis of each component. This is important in a prebiotic context because the early Earth was harsh and "hard" reactions were not likely to arise *de novo*.[29] Second, by definition, networks are highly cooperative interactions. Although a common view of the origin of life focuses on the advent of single self-replicating species, new evidence suggests that cooperation among molecular elements is not only possible but can grant fitness benefits to individual species.[30] And third, networks by definition distribute their composite function over several members, imbuing

[a] *Laboratoire de Biochimie, CNRS – ESPCI ParisTech, France*
[b] *SmartAnalytiX.com, Lausanne, Switzerland*
[c] *Institute for Systems Biology, Seattle, WA 98109, USA*
[d] *School of Earth and Space Exploration and Beyond Center for Fundamental Concepts in Science, Arizona State University, Tempe, AZ 85287, USA*
[e] *Department of Biochemistry, University of Missouri, 117 Schweitzer Hall, Columbia, MO 65211, USA*
[f] *Department of Chemistry, Portland State University, P.O. Box 751, Portland, OR 97207, USA. E-mail: niles@pdx.edu; Tel: +1-503-725-8769*
† These authors contributed equally to this work.

them with an buffering quality that allows them to lose individuals and yet persist even in fluctuating environments.[31] The prebiotic milieu was chaotic by all accounts, and thus having resilience from its very origins would have been an indispensible feature of life.

Despite the potential of networks in illuminating life's origins, we have not yet been able to realize their full promise in that role. This is in large part because, though networks have been extensively discussed from a theoretical point of view, two important aspects of their function in the origin of life are severely underexamined. The first is a concrete discussion of the chemical and physical realities of such networks and how the environment would have been key in fashioning the advent and propagation of molecular networks that led eventually to life. The second is how such networks could have evolved: what critical parameters govern selective forces in networks, keeping in consideration the first aspect. In this paper, we examine specifically the evolvability of prebiotic networks with an eye to plausible chemistry. As a result, our intent is to make network evolution a prebiotic plausibility and set the stage for empirical studies in the laboratory that can support, refine, or refute our conclusions.

What is chemical evolution?

As early as Miller[3] and Oparin,[4] it had been realized that chemical evolution must have preceded biological evolution. Chemical evolution entailed the heritable alteration of the identities of a collection of molecular species prior to formal genotype–phenotype relationships. As such, it must have had some capacity to encode information (*i.e.*, some reduction in ambiguity about all possible molecular ensembles), but clearly without the coding specificity available to biological systems. Chemical evolutionary settings have been described as "pre-life" in which information can be generated through selection in the absence of formal replication.[32] In this model, monomers become activated and polymerize into polymers with certain transition probabilities, and the dynamics of the system can be governed by a set of differential equations that convey the relative frequencies of polymer strings. This approach is very productive, and can foreshadow key evolutionary phenomena – such as selection and the existence of an error threshold – even without self-replication.

Yet the pre-life construct and others[33,34] often focus on the notion of polymers competing with each other for dominance. Should chemical cooperation among information-bearing molecules be required to crystallize replication and fend off parasitic side reactions,[30] then we need to consider the dynamics of network evolution as a priority. Jain & Krishna[35] provided a mathematical model and simulation study that has been perhaps the most direct analysis of this scenario to date. The model considered the reaction dynamics of a system of interconnected nodes (undefined molecular species) that may be able to catalyze the ligation of reactant molecules to produce other catalytic species. In that study, the appearance of autocatalytic sets, though unpredictable, was shown to be inevitable. Furthermore, one of the more intriguing patterns that the authors discovered was a tendency for network members to go nearly extinct through catastrophic stochastic fluctuations in frequencies but then recover through the appearance of a nascent autocatalytic set.[35] Such dynamics imply that there may be relatively simple rules for the growth and evolution of molecular networks in the absence of Darwinian-type replicator evolution. This is the sort of chemical evolution we envisage as preceding biological evolution, yet the vague nature of this model *vis-à-vis* known prebiotic chemistry needs to be clarified using more realistic abiochemistry and replicating species.

What are the features of a molecular network?

Networks are collections of interconnected entities: proteins that recognize other proteins, neurons that connect to other neurons, *etc.* In the language of graph theory they are described as nodes (entities) and edges (connections) (Fig. 1a; Table 1). For example, the yeast protein interaction network – its interactome – is composed of protein nodes that interact with each other, with the interactions represented as edges. In what follows, we are specifically concerned with what we call catalysis graphs. A catalysis graph is a network in which the nodes represent molecular species, and the edges represent catalytic interactions. In particular, if molecular species i catalyzes a reaction that produces molecular species j, then there is a (directed) edge from node i to node j. Examples of such catalysis graphs are presented in Fig. 2. Alternatively, the catalysis graph can be defined in terms of chemical reactions, where the nodes represent reactions and there is a (directed) edge from node i to node j if reaction i produces a molecule that can catalyze reaction j. For most practical purposes, and for our discussions below, these two representations can be considered equivalent. However it should be noted that they are not necessarily formally equivalent in all cases.[36]

The number of edges to which a node is connected is the node's degree. Moreover, in any network there will be cores, which are subsets of nodes in which every node is reachable from any other, *i.e.*, any pair of nodes is connected by a sequence of edges (a path) in each direction. In the language of graph theory, cores are strongly connected components (SCC) in the network. Cores can be as simple as a single loop of interconnected nodes. A core can also be so encompassing that any other species within the graph is not strongly connected to it, meaning that such outliers are either "upstream" or "downstream" of this core. Such cores are termed maximal cores. Finally, the maximal core that displays the highest net growth rate is termed the dominant core.

For networks to evolve, *i.e.*, change the frequencies and identities of their composite members as a function of time, there must be some means to describe explicitly how fitness can be realized in the context of a network. Unlike a typical biological scenario, this fitness can be manifest both within and between networks. This is because networks can overlap;

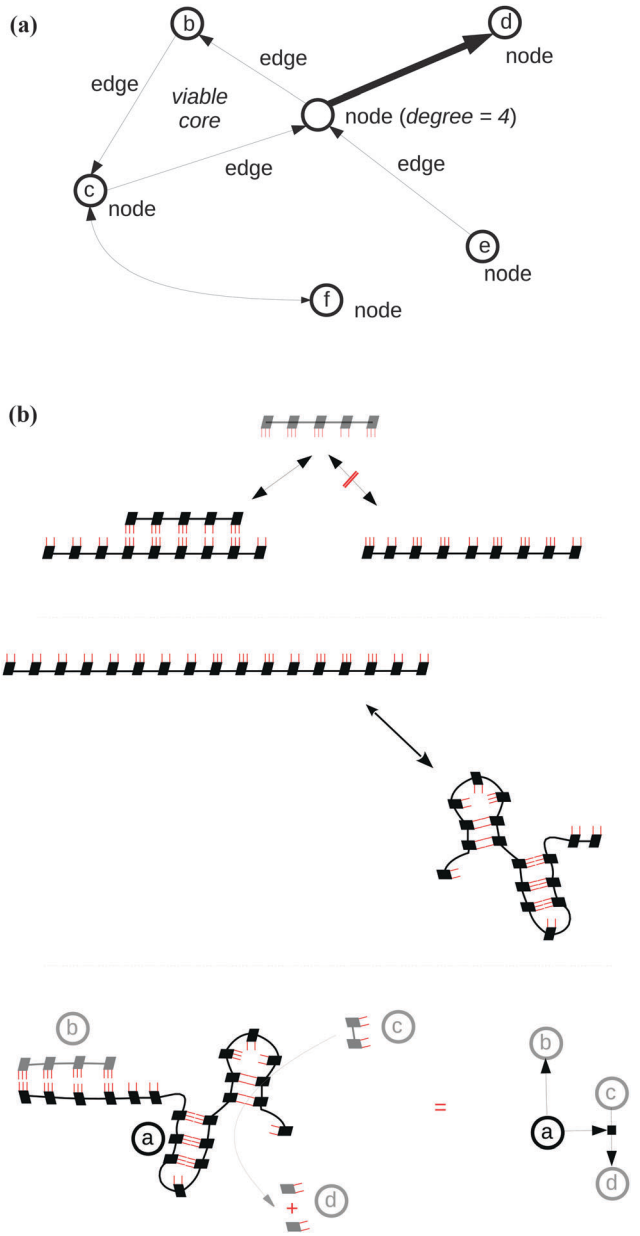

Fig. 1 Networks in prebiotic chemistry. (a) Example simple network among six molecular species. Here, each circle is a node, which represents a distinct molecule. Arrows are edges between nodes, and represent chemical transformations that can occur between molecular species. In this diagram, there is no catalysis or reactions among two or more molecules, only 1st-order interconversions. However the strengths (*i.e.*, rate constants) of the reactions can be depicted by the thickness of the arrow. Nodes c and f can interconvert in either direction. Nodes a, b, and c form a viable core, which is a closed set, and represents the smallest self-sustaining sub-network. (b) Three consequences of weak (*e.g.*, hydrogen) bonds in prebiotic networks. Top: A sensing function by polymers for other polymers. Middle: A complex structure formation. Bottom: An iteration function: multiple edges connected to a single node. In a molecular setting this could be a molecule, a, which is simultaneously (or sequentially) binding one substrate, b, while catalyzing the conversion of another, c to d, as depicted.

members of one network can also be members of others. In an evolutionary setting, successful, or fit, networks are those that persist and grow over time. We posit that this growth can be measured in terms of the proportion of polymers in a given environment that participate in a particular network, which in turn relates to the number and density of nodes within a network. However one can also assess the number of distinct environments into which a network has articulated itself. Both of these measures relates to the success of a network, and hence are reflections of its evolutionary past. Networks would compete with each other in a group selection process and perhaps a "quasi-species"[16,17] of self-sustaining autocatalytic sets is a plausible unit of selection.

Table 1 Terminology for networks and their evolution

Concept	Definition
Node	An information-bearing entity (*e.g.*, a molecular species), or input (food molecule) in a network
Edge	A connection between two nodes (either an intermolecular transformation or a catalytic interaction)
Degree	The number of nodes to which a node is connected (k)
Path length	The number of edges between two nodes (l)
Core	A set of nodes in which every node is reachable from any other
Viable core	A core that is also self-sustaining (given a suitable food source)
Maximal core	A core that cannot be extended by the inclusion of any additional existing node
Dominant core	A core that displays the highest net growth rate
Evanescent set	Reactions upstream of the dominant core that can fade from relevance with time
Hub	A node that is highly connected to other nodes
Autocatalytic set (CAS)	A collection of molecules that catalyze the necessary chemical reactions to synthesize all members of the set from an appropriate food source
RAF	A reflexively autocatalytic and food-generated set, *i.e.*, the mathematical formalization of an autocatalytic set
irrRAF	An irreducible RAF set, *i.e.*, a RAF set that does not contain a proper subset which is also RAF
Upstream	Specifies nodes outside of a core but that could feed into a core
Downstream	Specifies nodes outside of a core but could be produced by the core

Molecular networks from the perspective of known chemistry

Let us now consider the features of candidate molecules in prebiotic networks. Given an assumption that these are information-bearing polymers, then the most often discussed are polypeptides, lipids, or nucleic acids and their derivatives. The common feature that all three of these classes of molecules possess is the ability to form multiple non-covalent bonding interactions with other molecules of their own class. For peptides, this means hydrogen bonding, electrostatic interactions, and hydrophobic interactions among informational motifs (amino acid side chains). For lipids, this means hydrophobic interactions among structural motifs (aliphatic chains). For nucleic-acid-like molecules, this means hydrogen bonding between informational motifs (nucleobase edges)[37] and hydrophobic stacking among informational motifs (*e.g.*, π–π interactions between nucleobase planar surfaces). Of course, molecules of all three types can

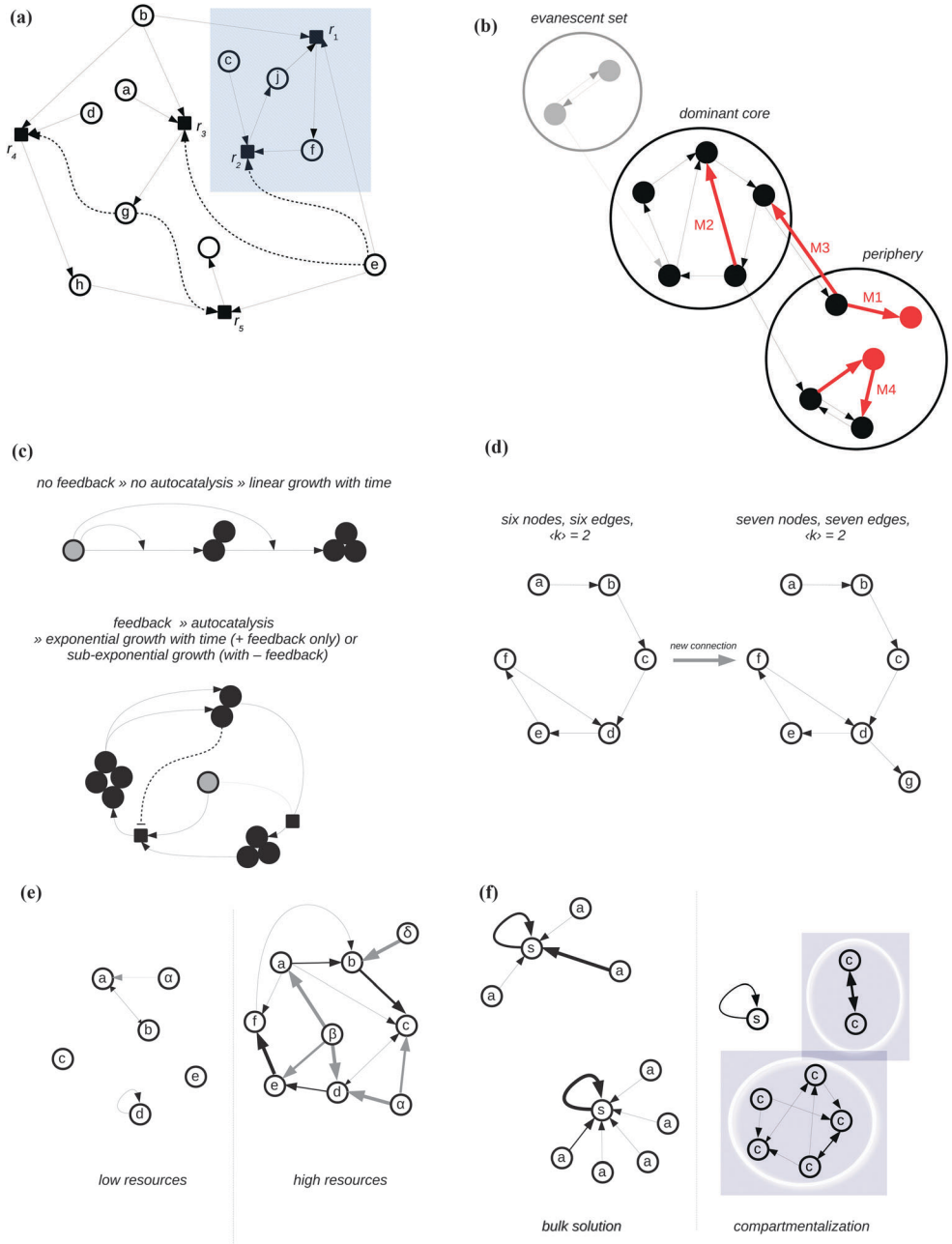

Fig. 2 Six key phenomena that govern prebiotic network evolution. (a) Viable cores. An example of a viable core – a catalytically closed and self-sustaining reaction network – in a simple polymer model where molecules are represented by circles, as in Fig. 1a. However, here boxes represent reactions among two or more of these molecules.[43] Solid arrows indicate reactants going into and products coming out of a reaction, while dashed arrows indicate which molecules catalyze which reactions. The food set consists of the species a–e. This particular autocatalytic set consists of five reactions (denoted r_1–r_5), and can be decomposed into two autocatalytic subsets {r_1, r_2} and {r_3, r_4, r_5}, the first of which (in shaded box) is an irreducible autocatalytic set, *i.e.*, a viable core. (b) Connectivity and kinetics. Prebiotic networks can expand by four modes (M1–M4; red arrows) within a dominant set and its upstream and downstream sets. A new node can be added in either the periphery (M1) or in the dominant core (M2) without affecting existing connections. Alternatively the dominant core can either expand by assimilating reactions in the periphery (M3) or shift to a new subset (M4). (c) Information control. The addition of negative feedback (dashed line) is critical for the establishment of stable networks. Figure adapted from ref. 92. (d) Scalability by preferential attachment.[61] The random network on the left has six nodes and a total of six edges, distributed binomially. One node has k = 1, four have k = 2, and one has k = 3. The exact nature of the nodes is undetermined: they could be either single chemical species, sets of redundant ones, cores, or networks. On the right, the network has grown by addition of one new node, 7, added to the previously existing node of highest degree. The average degree remains unchanged at 2.0 by this process; however, the degree distribution deviates from binomial. Four nodes of degree k = 2 remain. Two nodes, however, now have k = 1, and one has k = 4. Although the average degree of the network, $\langle k \rangle$, remains equal to 2.0, the distribution has become skewed toward nodes of lower and higher degree, eventually leading recursively to a scale-free topology, as shown in Fig. 3. (e) Resource availability. Food molecules (Greek letters) in prebiotic information networks can be non-information bearing compounds (*e.g.*, sugars to feed the synthesis of nucleotides) that can support (gray arrows) the maintenance, growth, and evolvability of networks both by driving preferential attachment[42,43,70] and by serving as control agents: minority molecules.[73] (f) Compartmentalization. The existence of a chemical or physical boundary within prebiotic networks can promote cooperation by excluding selfish replicators (s) that otherwise grow at the expense of altruists (a); this phenomenon is well established theoretically.[15,30,35,42]

interact non-covalently with members of other classes, and this may have been prebiotically important, but for simplicity we will only consider intra-class interactions here.

These weak bonds influence network formation in three discrete ways (Fig. 1b). First, low energy bonds can form and break rapidly, engendering a sensing function by polymers for other polymers. Second, they allow complex structure formation, such as secondary, tertiary, and quaternary contacts. And third, they allow for an iteration function in that multiple simultaneous or sequential interactions can form both in time and space, *e.g.*, multiple edges connected to a single node in the reaction graph. The combined effect of all three of these traits means that such polymers can interact with one another dynamically, iteratively, and through more than one type of mechanism. For example, a small coalition of polypeptides can transiently bind with each other *via* specific recognition rules determined by the identities of their composite amino acids (sensing). They concomitantly form catalytic structures that operate on other members of the coalition (structure formation). And they also simultaneously move from one interaction to another in a dynamic and measurable fashion (iteration). If the peptides possess certain subsets of amino acid sequences, then the net result is that autocatalytic sets can be established, as shown experimentally by Ghadiri and co-workers.[38,39] Importantly, without all three of these traits, either networks would not be able to form, or if they did, they would have very different characteristics. This can be seen in the case of sugar interconversions as in the formose reaction. Monosaccharides and their composite units such as formaldehyde and glyceraldehyde lack (to a large extent) sensing and complex structure formation. Although they can form catalytic networks, informational cohesion is not facile and the networks tend to expand in a chaotic fashion as time progresses.[26]

With these three traits, the critical role of the environment becomes apparent. The degree to which non-covalent interactions form within and among polymers is particularly dependent in a predictable fashion on polymer concentration, temperature, and local ionic conditions. Networks should be favored under conditions where polymer concentrations are high enough to allow repeated encounters with other polymers but not so high as to promote uncontrolled aggregation. The same is true of temperature and ionic (pH and salt) conditions. While there will be a broad region of parameter space that would satisfy these conditions, certainly there will be "goldilocks" regions where network formation – and evolution – will be the most productive. For single-stranded nucleic acids, for instance, concentrations in the low micromolar range at pH 5–8 should promote interactions frequent enough and yet strong enough to engender simple catalytic networks.[17,40,41] The key point is that for a given polymer, one can anticipate how environmental changes will modulate network dynamics.

Six key parameters for prebiotic network evolution

In their groundbreaking simulation-based study of evolution in autocatalytic sets, Vasas *et al.*[42] considered what exactly (Darwinian) evolvability means in a network setting. They agreed with previous authors[15,22] that alternative networks can coexist in the same environment, and proposed certain features, such as subsets that are "attractors" for reaction pathways, as being a requisite for evolution. Vasas *et al.* concluded that compartmentalization was an absolute requirement for evolvability. Taking those results as a foundation, and by considering a broad, eclectic swath of studies on network dynamics that exist in the literature – and by incorporating the chemical realities of likely polymer candidates – we have arrived at six key parameters that will determine how networks evolve, in a broader sense, once formed (Fig. 2).

1. Viable cores

A viable core[42] is a core (a strongly connected component of a network) in a chemical reaction system's catalysis graph that is also self-sustaining. In other words, a viable core is a collection of molecular species that mutually catalyze each other's formation. This property is termed catalytic closure. But to be a viable core a second requirement exists: that all relevant components can be constructed through sequences of such mutually catalyzed reactions starting from a suitable food source; *i.e.*, it must be self-sustaining. The food source is a collection of molecular species that can be assumed to be directly available from the environment. In many prebiotic scenarios of polymerization events, this food source would be amino acid or nucleotide monomers, for example. This notion of a catalytically closed and self-sustaining subset of molecules and reactions is formalized mathematically in terms of reflexively autocatalytic and food-generated (RAF) sets,[22] and efficient computer algorithms exist to detect and analyze such RAF sets in arbitrary[44] and realized[45] reaction networks.

It was previously shown that autocatalytic (RAF) sets can often be decomposed into smaller subsets which themselves are autocatalytic, in an iterative fashion until the smallest subsets (*i.e.*, the irreducible RAFs, or irrRAFs) are reached.[43] These subsets thus form a hierarchical structure or, more formally, a partially ordered set (POSET), with the union of all possible autocatalytic subsets at the top, and the individual irrRAFs at the bottom (Fig. 2a). Viable cores (*i.e.*, autocatalytic subsets) are essentially the primary unit of heredity in a network setting;[42] the eventual transition from network-based evolution to individual-based evolution may have mirrored the transition from cores to genes as the target for selection.

In principle there can be an exponentially large number of irrRAFs (or viable cores) within any given autocatalytic set.[43] In other words, it is possible to construct autocatalytic sets with n reactions that contain on the order of 2^n irrRAFs. In practice there also seem to exist large numbers of irrRAFs (possibly even thousands, although with a certain amount of overlap) in random instances of a simple polymer model, even with a maximum polymer length of only 10, and also in empirical systems[44,45] (*infra vide*). Thus, one of the main conditions for evolvability of autocatalytic sets as outlined in Vasas *et al.*,[42] namely the existence of multiple viable cores in the underlying reaction network on which selection can act, seems to be satisfied in several well-studied systems, both theoretical and

empirical. Recently a basic but formal example was provided of how this condition is satisfied in a simple reaction model, and how the possible viable cores (irrRAFs) can exist in different combinations inside compartments, thus potentially giving rise to different "phenotypes" and competition among them.[46]

An autocatalytic set will almost always be a system governed by non-linear dynamics and can display metastability. In fact, one common result of theoretical analyses is that such sets will often either settle into one or more oscillatory attractors (*i.e.*, steady states) each of which "drains" a basin of attraction states that flow to that attractor.[21,35] In other words, we would expect some collections of prebiotic molecules to assimilate other nearby collections and grow through accretion. We are now in a position to emphasize that cores can expand and contract by topological constraints (see parameter #2 below), and that having a core should lead networks to be scale-free (see parameter #4 below).

2. Connectivity kinetics

It is clear from the previous section that autocatalytic networks are characterized by their cores. However, it is not immediately obvious how the growth of the participating species is determined by connectivity. Which nodes are connected to which others – and the strengths of such connections – can impact the ability of a network to add or subtract members and hence change with time (Fig. 2b). We can examine the minimum requirements for such an analysis when the production rate of every species has a first-order dependence on the concentrations of upstream catalysts.[31] For now, we confine our consideration to positive catalytic links in the absence of resource limitations, these influences being further discussed below for parameters #3 and #5.

Note that a RAF set, as described above, necessarily contains one or more cores in the catalysis graph, but that such a core is not always a RAF set itself. For example, a core might not be self-sustaining, that is, might not be food-generated.[47] However, in the original model of Jain & Krishna[35] it is implicitly assumed that the relevant molecular species are directly produced from an unspecified food source. In other words, in such a simplified model the food-generated part of the definition of a RAF set is always (trivially) satisfied. As such, autocatalytic (RAF) sets and cores are equivalent, and thus any core is automatically also a viable core.

In this context, the Perron–Frobenius theory of linear algebra, applied to network dynamics, ensures the existence of a well defined steady exponential growth state that is governed by the structure of its maximal cores,[48] which we define here as the most inclusive subsets of species in which every pair is connected by a forward and backward reaction path. The growth rate of such a core taken in isolation, its intrinsic growth rate, is determined by a certain measure of cooperativity, and relates to the concept of "eigenvector centrality".[49] Precisely, for a length l, say $l = 3$, count the number of alternative catalytic paths of length 3 ($=a_3$) leading to a given species, then take its average over all species $\langle a_3 \rangle$. The value $\sqrt[l]{\langle a_l \rangle}$ converges rapidly to the exact growth rate with increasing l (ref. 50). For weighted networks, this scheme can be intuitively pictured by replacing a link of weight W by W parallel links of weight 1 (ref. 51).

Within an isolated reaction vessel, or "warm little pond", this model shows that a network possesses three categories of members: (i) those that participate in the core with the highest intrinsic growth rate, the dominant core, whose connectivity determines the overall growth rate; (ii) species catalyzed downstream of the dominant core, *i.e.*, those in the periphery, that inherits the growth rate of the dominant core; and (iii) structures upstream of the dominant core, the evanescent set, which tends to be exponentially negligible in the long run (Fig. 2b). Together, (i) and (ii) form an autocatalytic set (*i.e.*, a RAF set, given the caveat mentioned above that the food-generated part is trivially satisfied). Its species all grow at a same rate, but do so in different proportions determined by the connectivity and the strength of the catalytic links[52] overall defining a compositional identity of the network. These links may vary as the environment changes (see parameters #5 and #6), thus leading to a rudimentary form of chemical evolution. For example, temperature fluctuations would affect the K_D values of any polymers that interact through hydrogen bonding, and cause variations in the compositional identity of the network. Food-set variations and internal random events can lead to even larger-scale events such as random core shifts.[35]

In a more elaborate pre-biotic set-up, any form of compartment dynamics (see parameter #6 below) would induce competition between networks. Compartments that possess the fastest growing networks, *i.e.*, those which possess the most cooperative dominant core, would take over the population. Evolution could proceed in a quasi-Darwinian way, by stochastic and heritable transitions between networks.[42] At a certain evolutionary stage, potentially before generalist templated replication, networks may have developed the ability to acquire novel activities while maintaining extant useful ones. In this case, the classification above implies that only a few elementary modes of network expansion are compatible with network growth based on competition (red arrows in Fig. 2b). In Fig. 2b, M1 is periphery expansion: a new link appears in the periphery, without modifying the connectivity of the dominant core. This mode is neutral in terms of growth rate, but can prepare further evolution according to modes M3 or M4. M2 is core enhancement: new links are added between species within the dominant core. This leads to a growth rate increase by increasing the number of alternative paths. M3 is core expansion: a feedback is created from the periphery to the dominant core. The growth rate increase is then a consequence of an assimilation of new species in the dominant core that increase the overall cooperativity. And M4 is core shift: new links are added between species of the periphery and give rise to a new dominant core. The former dominant core is then by definition upstream of the new dominant core, and thus it is evanescent and disappears.

3. Information control

The limited availability of biological building blocks on the primitive Earth has led to a significant amount of work on feed-forward reaction networks in prebiotic chemistry, where the

express goal is to obtain as high a yield as possible of a desired biopolymer. Much less studied are inhibition, degradation, and other negative feedback control mechanisms. However, negative feedback plays an important role in biological robustness and evolution[52] and may have played an equally prominent role in the origins of life. Networks with strictly feed-forward network architecture are unstable. A prebiotically relevant example could be the hypercycle – cooperative sets of molecules where each molecule can self-replicate and catalyze the replication of another member of the same set.[17] Hypercycles are well-known to undergo oscillatory behavior in the population size of replicators for sets of four or more cooperative catalysts. This instability leads hypercycles to be vulnerable to perturbations including mutation in members of the hypercycle network and parasitism by other replicators. Despite the fact that each member of the hypercycle is capable of self-replication, these instabilities can lead to population collapse and extinction of the entire set. Perhaps as a consequence, an empirical demonstration of a strict hypercycle in the laboratory is so far lacking.

A possible resolution to the instabilities that arise in network expansion and evolution is to balance positive feedback, which enables growth, with negative feedback, which contributes to robustness and stability (Fig. 2c). In prebiotic systems with a limited supply of resources negative feedback may also be essential to selection.[53–56] Our study of recycling in a system of recombining RNA fragments demonstrated that negative feedback could provide a robust selection mechanism for prebiotic replicators.[54] In the study, simulations demonstrated that selection can favor a ribozyme that catalyzes degradation in systems with a limited supply of resources by selectively degrading unfavorable reaction products. This suggests that the catalytic landscape of RNAs – which is dominated by ribozymes that catalyze bond-cleaving reactions – could in fact play a beneficial role in early evolution under conditions of limited resource availability likely present on early Earth. Clearly negative feedback is critical in maintaining the operation of contemporary biology networks; the ubiquination pathway in protein regulation stands out as a good parallel. Thus, it is perhaps not paradoxical but an essential facet of the success of chemical evolution that many primitive catalysts perform degradative reactions.

Negative feedback may also play an important role in establishing a separation of timescales and the emergence of ''memory'' in chemical networks. The hierarchical structures observed in topological studies of autocatalytic networks should be characterized by different dynamical timescales for nested structural components.[57,58] Within such hierarchies, slower components can act as memory for construction of the entire network. Separation of timescales may have driven the emergence of the first hereditary units in prebiotic networks.[15] This is supported by computational simulations performed by Kaneko and colleagues,[59] which demonstrated that recursive reproduction of protocells enclosing catalytic networks relied on the presence of a ''minority molecule'' that is produced on a much slower timescale than other molecules in the network. Minority molecules, or slow components in a network, can dynamically acquire the role of hereditary information carriers, controlling reproduction of the entire network. These heredity units could be either template replicators or viable cores depending on the network architecture. In fact the notion of a ''genotype'' in the case of evolving networks is perhaps premature. The networks we describe here dispatch sequence information over several cooperative entities, a situation that is intermediate between the compositional genomes of abstract chemical reaction networks and templated replication. Cooperative information bearing molecules thus likely allowed the transition from compositional dynamics to a purely sequence-encoded one.

Reverse reactions also promote autocatalytic sets. In the simple binary polymer model already mentioned above, each possible pair of a molecule (bit string) and a reaction has a probability p of being included in the catalysis set (*i.e.*, the corresponding edge is added to the catalysis graph). Consider bi-directional reactions in which a molecule catalyzes both the forward (ligation) and the reverse (cleavage) reaction of a polymer such as protein or RNA. In that case, for $n = 10$ (*i.e.*, maximum bit string length of 10), each molecule needs to catalyze 2.6 reactions on average to have about a 50% chance of having a RAF set in a random instance of the model. Now, if one takes only forward (ligation) reactions, an average of 4.3 reactions catalyzed per molecule is needed to find RAF sets with 50% probability, *i.e.*, about 1.6 times more than in the previous case. Thus using only ligation reactions (not cleavages) requires a higher level of catalysis to get the same probability of finding RAF sets. In other words, cleavage (recycling) reactions clearly provide an advantage,[54,55] one that can even be seen in the unbounded formose reaction where the cleavage of longer sugars is the source of autocatalytic dynamics.[60]

4. Scalability

The properties of some networks scale with the size of the networks. Yet many do not, so it is worth asking whether one should consider scalability in prebiotic networks. We posit that the nascent forms of life would have the greatest potential for niche invasion in Earth's history, and clearly it would be of interest to investigate what would be the topology of a prebiotic molecular network as it expands to include many biochemical components, regardless of the order in which those components were added.

Most, although not all, networks in Biology are sparse, small-world (or ultra-small-world), and scale-free. Sparseness refers to the fact that most species (*i.e.*, nodes) are not connected to each other. Small-world networks are those where the path between two connected nodes can be traversed through a limited number of edges. The scale-free property refers to the distribution of connections (*i.e.*, edges) among nodes. In any network, some nodes have many more connections (are of higher degree) than others; these would be hubs. A hub usually refers to a node in the network that happens to have a very high degree (*i.e.*, it is connected to many other nodes), but does not necessarily refer to a fully connected subset. A core by contrast is a connected sub-network such as in a catalytically closed RAF set; hubs refer to just single nodes.

The property of being scale-free has not only been observed in many existing networks of many kinds, but it has also been discovered as a property of growing under rather general mathematical conditions, including common versions of preferential attachment. The key point is that a scale-free network should exhibit a particular power law that relates the number of connections to a node (its degree, k) to the probability that a node has this degree (Fig. 3). The relationship $P(k) \sim k^{-\gamma}$ defines a scale-free network, and in fact such structures typically generate γ values near 2 (ref. 61). Scale-free networks contrast with random networks, often called either Erdös-Rényi – after the mathematicians who developed the model – or Poisson networks, because the degrees per node follow a Poisson distribution with a well-defined mean. By contrast, the term "mean degree" is not meaningful in discussing scale-free networks except in the trivial sense of $\sum k_i/N$, where N is the number of nodes and k_i the degree of each node. The most common mechanisms for scale-free networks to arise is in fact through growth and preferential attachment. In other words, higher-degree nodes are more likely to form new edges than are lower-degree nodes (Fig. 2d). Note that formally a network can only be scale-free as it approaches an infinite number of nodes;[62] here we use the term scale-free to approximate this behavior.

The distinction between Poisson and scale-free networks extends to differences in information content carried by their structures, which could be measured through the Shannon entropy or other measures.[63,64] Scale-free distributions have lower entropy than Poisson distributions. This conclusion has been derived by statistical mechanical calculations of the ensemble of possible networks;[63] however, it is also possible to be visualized non-mathematically. In the preferential attachment model for the generation of a scale-free network,[61] new nodes that are added to the network are more likely to add to pre-existing nodes (hubs) of higher degree (Fig. 2d). Therefore the number of choices available to a new node is lower in a scale-free network than in one where all nodes are equivalent – by definition a situation of lower entropy.

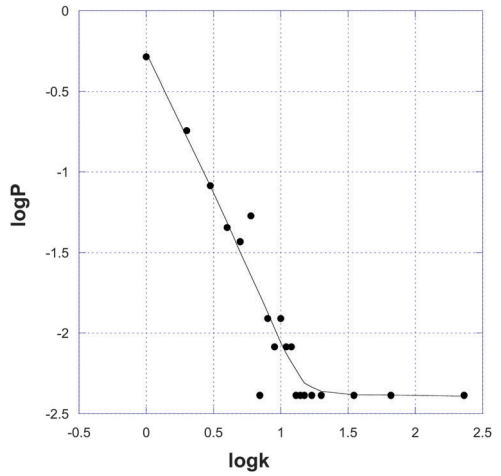

Fig. 3 Log–log plot of relative occurrence $P(k)$ of RNA hairpin motifs *vs.* the degree k of times they appear in the nonredundant RNA structural database (http://rna.bgsu.edu/rna3dhub/nrlist/).

In fact, there is some evidence that autocatalytic (RAF) sets could form even more easily in reaction networks with a power law (rather than a Poisson) distribution in the degree of catalytic connections.[47]

What chemistry permits lower entropy in forming a network? We suggest three possible mechanisms that have experimental precedent in evolving from a random to a scale-free topology. In this case, there must be an advantage to the nodes that become hubs. This advantage need not be intrinsic to the hub species itself and may instead originate from connectivity dynamics. One possible advantage to a pre-hub node would be the luck of the draw. Any species in a Poisson network that by chance has a degree higher than the mean has an advantage when selection follows a preferential attachment dynamic, as in a periphery expansion described above. In a prebiotic setting, if networks grow by preferential attachment, a concrete consequence of these mechanisms could be the inevitability of scale-free networks.[61] Consider, for example, the origin of templated replication, a process that must have arisen at some point given its ubiquity in biology. Any random polymerization process when followed by a templated process will be captured by selection because templating can give benefits to those sequences that could successfully direct it.[10,30] One can then conclude that selection for higher connectivity leads to greater network robustness in that a certain degree of sequence similarity across nodes is favored.

A second means of preferential attachment could come through recombination that is directed by limited homology.[65–67] In this model, a node sequence or its complement that was over-represented in a random network would be able to grow by addition of other sequences, eventually leading to a template-competent sequence. In other words, short polymers that are more promiscuous in their interaction partners, for example as a consequence of "wobble"-like hydrogen-bond pairing, might emerge as hubs. Thus assortment based on the recognition of a short random sequence will lead to a scale-free network. While this may be clear in the case of RNA networks, the weak bonds between nodes in protein or lipid networks, considering too that such molecules can also recombine, means that this should be a general phenomenon.[27,28]

A third mechanism for entropic change is criticality.[68] Critical networks have power law distribution of "avalanches of change" when the dynamics of one variable is altered. Such critical behavior, found in neural systems and cell genetic regulatory networks, may optimize the capacity to categorize the environment (given inputs from it) and respond reliably without chaotic behavior. For example, a 9-peptide autocatalytic set has been demonstrated in which all nodes can realize any Boolean function on its $k = 2$ inputs.[39] In fact, binary variable networks, with $k = 2$, with a random choice of Boolean function for each node, and with random sets of interconnections, are generically critical. For k greater than 2, a bias in the ratio of values in the Boolean function – $P(k)$ and k (for Erdös-Rényi networks), or $P(k)$ and γ (for scale-free networks of power law slope γ) – has a one-dimensional line of criticality separating order from high dimensional chaos. Moreover it has been shown

that critical nets minimize alteration in attractors when connections (or logic) are altered slightly, and thus evolve "gracefully", and has shown that if one starts with chaotic or ordered Boolean networks and selects over generations only for minimizing changes to attractors, the population evolves to criticality.[69] Even if a network is Erdös-Rényi in structure at the start, it likely becomes scale-free eventually.

5. Resource availability

Because information polymers are in fact polymers, their synthesis – prebiotic or biotic – will require the availability of "food" molecules, typically monomers. The density and availability of such molecules will determine the evolvability of a network. In systems with a finite supply of resources, recycling rates will also be important for evolvability.[54,70] As described above, the RAF concept explicitly accommodates the interplay between food molecules and the polymers that they support. This interaction, often overlooked, is an important consideration for network stability. In fact, it can allow for multiple stable states; for example, food-generated sets spontaneously colonize the space of successively catalytic species, but rare random events that lead to open access to non-food generated sets are typical events that trigger heritable novelties and jumps in complexity.[43]

Importantly, the evolvability of networks differs between food-rich scenarios and those in which resources are limiting (Fig. 2e). There can be common resources for the different members of a given network, so that a balance appears between cooperativity (parameter #2 above) and resource titration; this could favor certain cores. Also the diversity of the food has been shown to be important. For example, consider the random polymerization of small oligonucleotides. Here their flux may be biased toward some types of sequences.[32,71] Moreover upstream phenomena in the autocatalytic network can be considered as being a "proto-metabolism" that pre-processes input chemicals. At some point, the autocatalytic set could incorporate feedback (positive or negative, see parameter #3 above) by catalyzing upstream events and integrating the proto-metabolism. Finally it should be noted that an additional necessary feature for the selection of replicators with resource-dependent replication rates is that the replicators – and the environment they are in – first synchronize their composition.[72]

6. Compartmentalization

The advantages of compartments to primordial networks were emphasized by Vasas *et al.*[42] Cell-like structures could be protocells with some type of polymeric boundary, lipid vesicles, or more rigid compartments such as rock fissures. They also may display many different types of dynamics, from unstable compartments with large pooling events and reformation of smaller compartments (think small ponds on an uneven surface), to a mother-daughter situation (think membrane-bound spheres swelling and budding). In fact, the dynamics of minority molecules mentioned above[59,73] can induce spatial clustering, so that the compartment arises as a consequence of information control in the reaction network structures.[74]

Compartmentalization can also refer to phase partitioning, as realized on solid surfaces at a solid–liquid interface as noted by Bernal,[2] or in aerosols at the air–water interphase as proposed by Woese.[74] For an evolving network, physico-chemical adsorption may enhance the concentration of marginal chemical species. This phenomenon has been noted as an enhancement for abiotic polymerization reactions,[33,75] and molecular crowding can speed up the rate of RNA-directed catalysis.[33,76] As a consequence, rare or minority members of a network can be recruited more easily, thereby affecting the connectivity kinetics (parameter #2 above) and information control (parameter #3 above). Unless phase partitioning results in a complete barrier to the supply of "food" molecules, our prediction would be that network growth and diversity would be promoted in these scenarios.

In any format, compartmentalization provides protection for networks against parasitic reactions that can prevent growth and adaptation, and they provide the potential for inter-compartmental competition that can drive evolution (Fig. 2f). Here we reiterate these advantages, but also posit that networks may form and evolve in the absence of compartments. For this to occur, new members and/or changes in strengths of connections among nodes must take place, as we describe especially under parameter #2 above. At first glance such alterations seem impossible if the networks are not physically separated from each other. Yet one can imagine a collection of polymers that are forming a network that is determined by the weak bonding interactions that persist under the prevailing environmental conditions. If these are peptides, for example, then the K_D values that determine which peptides interact through ionic interactions among amino acid side chains would be strongly dependent on salt concentrations or pH. Should the environment suddenly change through the increase of ionic strength as, say, a prebiotic pool evaporates, then new members (*i.e.*, periphery expansion) could be added to the network that were previously present in the milieu but not influential. At the same time the strengths of the connections within cores would be affected (core enhancement). With nucleic acids, temperature changes – and these would have been daily and severe in many primordial environments at 4 Gya – would affect T_m values between polymers, leading to the same sorts of phenomena. Evolution in this sense would be primarily kinetic in nature, and yet this type of change has been postulated on theoretical grounds[45,77,78] and seen in empirical studies of RNA networks in the laboratory.[66]

Setting the stage for empirical experiments

It should be apparent from the above descriptions that all of these parameters are highly interdependent. But can any of these patterns be tested empirically? As early as 1998 with the graded autocatalysis replication domain (GARD) model, Lancet and colleagues realized that lipids should form information-bearing networks,[79] and later simulations have shown that many systems in which mutual catalysis exists in excess over self-catalysis should be evolutionarily favored.[80,81] This result

has been challenged,[82] and without some means of translating such ideas into measurable phenomena, our understanding of chemical evolution is rather limited.

Some work has already been done to examine networks in this context. Peptide replicators can operate in a network format in which positive feedback exists.[38,39,83] The number of coexisting nodes with replicator peptides has recently been extended to six,[84] and there is evidence that diverse networks based on structure can form with peptides.[85] Networks have also been established with short tri-mer RNA fragments[86] and with up to 48 RNAs cooperating to form ribozymes.[66,87] There is also evidence that RNA can form scale-free networks, at least in contemporary Biology. We examined distribution of hairpin motifs in a non-redundant RNA structural database and found that most hairpin motifs are represented only 1–2 times in the database, while a few, especially the GNRA tetraloop, are represented many times more than the mean. The linear portion of a log–log plot of this relationship has negative slope $\gamma = 2$, consistent with scale-free topology (Fig. 3). While this contemporary distribution is clearly the result of natural selection, it should be possible, through parallel SELEX experiments, to determine the extent to which the node distribution represents an intrinsic property, *e.g.*, thermal stability of the favored tetraloops, or is simply a result of chance in evolution. For compartmentalized RNAs, something rarely discussed, but in practice very important, is the typical number of copies per compartment. For example in microfluidic droplets one needs $>10^5$ RNA to detect an activity, but most simulations only accommodate a handful ($\sim$10) of molecules. Copies/cell is a very important parameter for the transmission of information, for example regarding the size of the daughter material. Here enters the consideration of minority molecules that can ensure a better genotypic identity of compartments that otherwise need a large number of objects to have a well defined phenotype.

Many experiments have the potential to test network fitness, which will be a key parameter to quantify and examine moving forward. A comparison of node density and/or persistence among simple protein or RNA networks in the lab as a function of time would be a low-hanging fruit in this regard. Importantly, to assess the likelihood of prebiotic networks with monotypic polymer solutions (*e.g.*, all RNA), we would need to know the probability that random sequences encode functional polymers. Some progress toward this probability has been made. Using phage display and resistance to thrombin degradation, it was estimated that about 20% of random protein sequences of length 50 were folded.[88] However the density of specific functions in sequence space is most certainly far less. The frequency of GTP aptamers in the space of 24-mer RNAs was measured to be about 10^{-13} (ref. 89) and the 50% odds of hitting an isoleucine aptamer or a hammerhead ribozyme in the space of 100-mer RNAs was estimated at about 10^{-10} (ref. 90). The persistence of such nodes in an unstable simulated prebiotic environment could – and should – be examined by varying critical parameters such as the salt concentration, as discussed by Jiménez *et al.*[89] Such data should be compared to the wealth of network growth analyses that exist in other realms.[91]

With a knowledge of node density in random polymer space, specific network evolution experiments can be carried out to test specific predictions. Such experiments will require tracking both node and edge frequencies as they change over time under various selection pressures. One main prediction that we can make is that there is a process analogous to ecological succession in the evolution of networks.[66] "Weedy" sets such as irrRAFs, should form easily, but not be robust to environmental fluctuations. The addition of new nodes by a set of (as of yet not fully known) rules such as preferential attachment will then create more robust networks that are more resilient; these are capstone species in early chemical evolution.

Acknowledgements

SIW thanks the Templeton World Charity Foundation for funding support (opinions expressed in this publication are those of the authors and do not necessarily reflect the views of Templeton World Charity Foundation). NL would like to acknowledge NASA Exobiology and Evolutionary Biology NNX14-AK21G for funding.

References

1 J. B. S. Haldane, *The Rationalist Annual*, 1929, **148**, 3–10.
2 J. D. Bernal, *The Physical Basis of Life*, Routledge and Paul, London, UK, 1951.
3 S. L. Miller, *Science*, 1953, **117**, 528–529.
4 A. I. Oparin, *The Origin of Life on the Earth*, Academic Press, New York, NY, 1957.
5 J. von Neumann, *Theory of Self-reproducing Automata*, University of Illinois Press, Urbana, IL, 1966.
6 T. Gánti, *The Principle of Life (in Hungarian)*, Gondolat, Budapest, Hungary, 1971.
7 G. F. Joyce, *Nature*, 1989, **338**, 217–224.
8 J. Maynard Smith and E. Szathmáry, *The Major Transitions in Evolution*, Oxford University Press, Oxford, UK, 1995.
9 D. W. Deamer, *Microbiol. Mol. Biol. Rev.*, 1997, **61**, 230–261.
10 L. E. Orgel, *Nature*, 1994, **358**, 203–209.
11 T. Wu and L. E. Orgel, *J. Am. Chem. Soc.*, 1992, **114**, 317–322.
12 W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner and D. P. Bartel, *Science*, 2001, **292**, 1319–1325.
13 H. S. Zaher and P. J. Unrau, *RNA*, 2007, **13**, 1017–1026.
14 J. Attwater, A. Wochner and P. Holliger, *Nat. Chem.*, 2013, **5**, 1011–1018.
15 T. Gánti, *The Principles of Life*, Oxford University Press, Oxford, UK, 2003.
16 M. Eigen, *Naturwissenschaften*, 1971, **58**, 465–523.
17 M. Eigen and P. Schuster, *Naturwissenschaften*, 1977, **64**, 541–565.
18 M. Eigen and P. Schuster, *Naturwissenschaften*, 1978, **65**, 7–41.
19 M. Eigen and P. Schuster, *Naturwissenschaften*, 1978, **65**, 341–369.
20 S. A. Kauffman, *J. Cybernetics*, 1971, **1**, 71–96.

21 S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, Oxford, UK, 1993.
22 W. Hordijk and M. Steel, *J. Theor. Biol.*, 2004, **227**, 451–461.
23 G. Wächtershäuser, *Microbiol. Rev.*, 1988, **52**, 452–484.
24 C. De Duve, *Blueprint for a Cell*, Neil Patterson Publishers, 1991.
25 E. Branscomb and M. J. Russell, *Biochim. Biophys. Acta*, 2013, **1827**, 62–78.
26 H. Kuhn, *Angew. Chem., Int. Ed.*, 1972, **11**, 798–820.
27 L. E. Orgel, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 12503–12507.
28 S. A. Kauffman, *J. Theor. Biol.*, 1986, **119**, 1–24.
29 R. Shapiro, *Sci. Am.*, 2007, **296**, 46–53.
30 P. Higgs and N. Lehman, *Nat. Rev. Genet.*, 2015, **16**, 7–17.
31 C. N. Hinshelwood, *J. Chem. Soc.*, 1952, **1952**, 745–755.
32 M. Nowak and H. Ohtsuka, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 14924–14927.
33 C. Fernando, G. von Kiedrowski and E. Szathmáry, *J. Mol. Evol.*, 2007, **64**, 572–585.
34 M. Kreysing, L. Keil, S. Lanzmich and D. Braun, *Nat. Chem.*, 2015, **7**, 203–208.
35 S. Jain and S. Krishna, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 543–547.
36 M. Feinberg, *Chem. Eng. Sci.*, 1987, **42**, 2229–2268.
37 N. B. Leontis, J. Stombaugh and E. Westhof, *Nucleic Acids Res.*, 2002, **30**, 3497–3531.
38 D. H. Lee, K. Severin, Y. Yokobayashi and M. R. Ghadiri, *Nature*, 1997, **390**, 591–594.
39 G. Ashkenasy, R. Jagasia, M. Yadav and M. R. Ghadiri, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 10872–10877.
40 J. C. Striggles, M. B. Martin and F. J. Schmidt, *RNA*, 2006, **12**, 353–359.
41 S. Wu, Y. Chen, G. Mao, S. Trobro, M. Kwiatkowski and L. A. Kirsebom, *Nucleic Acids Res.*, 2014, **42**, 631–642.
42 V. Vasas, C. Fernando, S. Kauffman and E. Szathmáry, *Biol. Direct*, 2012, **7**, 1.
43 W. Hordijk, M. Steel and S. Kauffman, *Acta Biotheor.*, 2012, **60**, 379–392.
44 W. Hordijk, J. I. Smith and M. Steel, *Algorithms Mol. Biol.*, 2015, **10**, 15.
45 W. Hordijk and M. Steel, *J. Syst. Chem.*, 2013, **4**, 3.
46 W. Hordijk and M. Steel, *Origins Life Evol. Biospheres*, 2014, **44**, 111–124.
47 W. Hordijk, L. Hasenclever, J. Gao, D. Mincheva and J. Hein, *Nat. Comput.*, 2014, **13**, 287–296.
48 S. Jain and S. Krishna, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 2055–2060.
49 M. D. Konig and C. J. Tessone, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2011, **84**, 056108.
50 R. von Mises and H. Pollaczek-Geiringer, *Z. Angew. Math. Mech.*, 1929, **9**, 58–77.
51 M. E. J. Newman, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2004, **70**, 056131.
52 K. Klemm, M. A. Serrano, V. M. Eguíluz and M. San Miguel, *Sci. Rep.*, 2012, **2**, 292.
53 S. I. Walker and P. Davies, *J. R. Soc., Interface*, 2013, **6**, 20120869.
54 N. Vaiyda, S. I. Walker and N. Lehman, *Chem. Biol.*, 2013, **20**, 241–252.
55 G. A. M. King, *BioSystems*, 1982, **15**, 89–97.
56 G. A. M. King, *J. Theor. Biol.*, 1986, **123**, 493–991.
57 J. C. Flack, *Philos. Trans. R. Soc. London, Ser. B*, 2012, **367**, 1802–1810.
58 J. C. Flack, D. Erwin, T. Elliot and D. C. Krakauer, in *Evolution of Cooperation and Complexity*, ed. K. Sterelny, R. Joyce, B. Calcott and B. Fraser, MIT Press, Cambridge, MA, 2013, pp. 45–74.
59 K. Kaneko and T. Yomo, *J. Theor. Biol.*, 2002, **214**, 563–576.
60 N. Virgo and T. Ikegami, *Adv. Artif. Life ECAL*, 2013, **12**, 240–247.
61 A. L. Barabasi and R. Albert, *Science*, 1999, **286**, 509–512.
62 M. P. H. Stumpf, C. Wiuf and R. M. May, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 4221–4224.
63 G. Bianconi, *Europhys. Lett.*, 2008, **81**, 28005.
64 T. D. Schneider, *Nano Commun. Networks*, 2010, **1**, 173–180.
65 A. B. Chetverin, H. V. Chetverina, A. A. Demidenko and V. I. Ugarov, *Cell*, 1997, **88**, 503–513.
66 N. Vaidya, M. L. Manapat, I. A. Chen, R. Xulvi-Brunet, E. J. Hayden and N. Lehman, *Nature*, 2012, **491**, 72–77.
67 N. Lehman, C. Díaz Arenas, W. A. White and F. J. Schmidt, *Entropy*, 2011, **13**, 17–37.
68 G. Vattay, D. Salahub, I. Csabai, A. Nassimi and S. A. Kauffman, 2015, *arXiv*:1502.06880v1.
69 M. Aldena and P. Cluzel, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 8710–8714.
70 S. I. Walker, M. A. Grover and N. V. Hud, *PLoS One*, 2012, **7**, e34166.
71 J. Derr, M. L. Manapat, S. Rajamani, K. Leu, R. Xulvi-Brunet, I. Joseph, M. A. Nowak and I. A. Chen, *Nucleic Acids Res.*, 2012, **40**, 4711–4722.
72 C. Mathis, T. Bhattacharya and S. I. Walker, 2015, arXiv:1503.02776.
73 A. Kamimura and K. Kaneko, *Phys. Rev. Lett.*, 2010, **105**, 268103.
74 C. R. Woese, *J. Mol. Evol.*, 1979, **13**, 95–101.
75 J. Ferris, *Elements*, 2005, **1**, 145–149.
76 C. A. Strulson, R. C. Molden, C. D. Keating and P. C. Bevilacqua, *Nat. Chem.*, 2012, **4**, 941–946.
77 A. Pross, *Pure Appl. Chem.*, 2005, **77**, 1905–1921.
78 A. Pross, *J. Syst. Chem.*, 2011, **2**, 1.
79 D. Segre, D. Lancet, O. Kedem and Y. Pilpel, *Origins Life Evol. Biospheres*, 1998, **28**, 501–514.
80 D. Segre, D. Ben-Eli and D. Lancet, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 4112–4117.
81 M. Omer and D. Lancet, *Artif. Life*, 2012, **18**, 243–266.
82 V. Vasas, E. Szathmáry and M. Santos, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 1470–1475.
83 Z. Dadon, N. Wagner and G. Ashkenasy, *Angew. Chem., Int. Ed.*, 2008, **47**, 6128–6136.

84 Z. Dadon, N. Wagner, A. Alasibi, M. Samiappan, R. Mukherjee and G. Ashkenasy, *Chem. – Eur. J.*, 2015, **21**, 648–654.
85 W. S. Childers, N. R. Anthony, A. K. Mehta, K. M. Berland and D. G. Lynn, *Langmuir*, 2012, **28**, 6386–6395.
86 D. Sievers and G. von Kiedrowski, *Nature*, 1994, **369**, 221–224.
87 T. A. Lincoln and G. F. Joyce, *Science*, 2009, **323**, 1229–1232.
88 C. Chiarabelli, J. W. Vrijbloed, D. De Lucrezia, R. M. Thomas, P. Stano, F. Polticelli, T. Ottone, E. Papa and P. L. Luisi, *Chem. Biodiversity*, 2006, **3**, 840–859.
89 J. I. Jiménez, R. Xulvi-Brunet, G. W. Campbell, R. Turk-MacLeod and I. A. Chen, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 14984–14989.
90 R. Knight, H. De Sterck, R. Markel, S. Smit, A. Oshmyansky and M. Yarus, *Nucleic Acids Res.*, 2005, **33**, 5924–5935.
91 S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nodes to the Internet and WWW*, Oxford University Press, Oxford, UK, 2003.
92 R. Braakman and E. Smith, *Phys. Biol.*, 2013, **10**, 011001.

