Composing life

Daniel Segré and Doron Lancet+
Department of Molecular Genetics and The Crown Human Genome Center, The Weizmann Institute of Science, Rehovot 76100, Israel

Received June 15, 2000; revised July 28, 2000; accepted August 4, 2000

Textbooks often assert that life began with specialized complex molecules, such as RNA, that are capable of making their own copies. This scenario has serious difficulties, but an alternative has remained elusive. Recent research and computer simulations have suggested that the first steps toward life may not have involved biopolymers. Rather, noncovalent protocellular assemblies, generated by catalyzed recruitment of diverse amphiphilic and hydrophobic compounds, could have constituted the first systems capable of information storage, inheritance and selection. A complex chain of evolutionary events, yet to be deciphered, could then have led to the common ancestors of today’s free-living cells, and to the appearance of DNA, RNA and protein enzymes.

Planetary random chemistry
Specific functions in the living cell have traditionally been identified with specific classes of molecules. Yet the last few decades have seen numerous exceptions including RNA catalysis (Cech, 1993; Scott, 1998; Gesteland et al., 1999) and non-protein enzyme mimetics (Fendler, 1982; Vandersteen et al., 1996), as well as peptide- (Lee et al., 1997), lipid- (Bachmann et al., 1992; Kust and Rathman, 1995) and mineral-based (Cairns-Smith, 1982) self-replication. Also pyrite (Wachtershauser, 1988; Huber and Wachtershauser, 1997) and thioesters (de Duve, 1995; Weber, 1998) were proposed to have played a role in early energy metabolism. These relaxed boundaries of molecular function have profound implications for our understanding of the origin of life, including a potential prebiotic scenario whereby the functions of information storage, catalysis, energy transfer and compartmentalization could be bestowed upon a large variety of different chemical structures.

Nevertheless, the mainstream prebiotic evolutionary scenario, the RNA world (Gilbert, 1986; Sievers and von-Kiedrowski, 1994; Bolli et al., 1997; Gesteland et al., 1999), is based on a very narrow subset of chemicals. The roots of this hypothesis reside in the notion that ‘only a digital genetic system is capable of sustaining Darwinism over eons of geological time’ (Dawkins, 1996). While the RNA world view is supported by theory (Eigen, 1971; Eigen and Schuster, 1979; Kuppers, 1983) and experimentation (Fijalkowska and Schaaper, 1996; Ellington et al., 1997; Wright and Joyce, 1997; Gesteland et al., 1999), it encounters considerable difficulties when confronted with prebiotic Earth constraints (Shapiro, 1984, 2000; Yarus, 1999). How would local high concentrations of energized ribonucleotide monomers of the right kinds have formed? How can the formation and stability of the long RNA polymers needed for replication and catalysis be accounted for? How would the precious few sequences capable of replication ever appear?

An alternative appears to be necessary for the RNA-centric paradigm of the origin of life. While RNA must have had a central role at some point in cellular evolution, it may have not been the first and only player (Lahav, 1991; Segré and Lancet, 1999). Rather, life on our planet could have begun as a random chemistry melting pot, a ‘garbage-bag world’ (Dyson, 1999), with myriads of different chemical configurations (Morowitz, 1992; Segré and Lancet, 1999; Yarus, 1999). In a more restricted analogy, it has been proposed that the presently known RNA chemistry emerged from a large ‘gemisch’ of nucleotide analogs (Sievers and von-Kiedrowski, 1994; de Duve, 1995; Bolli et al., 1997; Wills and Bada, 2000). If so, then the chemistry of life as we know it today—polynucleotides, polynucleic acids and polysaccharides—has not been a fortuitous departure point, but the end result of elaborate selection and evolution. The crucial origin of life question then becomes how natural selection was initiated by some molecular assortments, irrespective of their exact chemistry (Lifson and Lifson, 1999).

Cytogenesis recapitulates biogenesis
During cell division, thousands of constituents, including enzymes, cytoskeletal components, organelles and membranes are duplicated. This biosynthetic ‘copying’ is made possible by a highly organized metabolic network (Ouzounis and Karp, 2000;
Schuster et al., 2000), and is followed by balanced allocation of molecular components, leading to the generation of two progeny. Interestingly, however, no known cellular constituent is capable of self-replication in pure form. Even DNA is absolutely dependent on other cellular components for making its own copies. Would it be reasonable to assume that prebiotic chemicals could outperform all present-day biomolecules? One is compelled to consider an alternative: that self-replication has never been a property of individual molecules, but rather one of molecular ensembles (Oparin, 1957; Dyson, 1982; Kauffman, 1986; Morowitz, 1996; Ganti, 1997; Dyson, 1999; Segré et al., 2000a). Perhaps at this early stage in biogenesis, the transfer of molecular content from a protocell to its progeny, similar to what happens in present-day cytogenesis, was the only inheritance possible.

Compositional genomes

A crucial problem for any scenario for the origin of life is the mechanism for information storage and propagation. For a molecular assembly devoid of genetic machinery, compositional information would be preserved only through a few divisions, eventually being lost to dilution. For information transfer to occur in perpetuity, a detailed mechanism of directed compound replenishment must exist. Homeostatic growth models for self-replication (Oparin, 1957; Kauffman, 1993; Ganti, 1997; Dyson, 1999; Segré et al., 2000a) account for this by linking increased size to chemical processes that preserve molecular concentrations (Wills and Bada, 2000). In Kauffman’s model, ‘catalytic closure’ is achieved through molecular interactions that gradually increase molecular diversity (Farmer et al., 1986; Kauffman, 1986, 1993). In Dyson’s scheme, highly organized catalytic interactions develop over time in a manner dependent on assembly size and the diversity of the components involved (Dyson, 1982, 1999). For Ganti’s self-reproducing fluid automaton (‘chemoton’), cycle stoichiometry appears within membrane-enclosed molecular assemblies (Ganti, 1975, 1997).

The schemes mentioned above provide mechanisms whereby an assembly might grow to twice its original size by absorbing and/or synthesizing additional molecules similar to those that it already contains. It will then be ready to split, generating two similar progeny, bearing what could be viewed as a ‘compositional genome’ (Segré and Lancet, 1999; Segré et al., 2000a) (Figure 1). Compositional genomes may harbor ‘unlimited heredity’ as previously defined (Szathmary and Maynard Smith, 1997). Importantly and in contrast to RNA-based models, compositional models involve practically no a priori constraints on molecular structure.

A lipid world

The chemical inventory on Earth in prebiotic times was all but meager (Leach et al., 1978; Clyba and Sagan, 1992; Matthews, 1992; Morowitz, 1992). Organic compounds are thought to have been available from many sources including interstellar space (Deamer, 1997; Allamandola et al., 1999), the atmosphere (Schlesinger and Miller, 1983), volcanoes (Pudkletnov and Markhinin, 1981) and hydrothermal vents (Lasaga, 1971; Amend and Shock, 1998; McCollom et al., 1999). It would thus be legitimate to dissociate the question of the origin of organics from that of the origin of life. The crucial enigma seems to be related to self-organization and the self-reproduction of supramolecular structures, and not to organosynthesis.

While self-assembly could involve substances that form colloidal coacervate particles (Oparin, 1957) or proteinoid microspheres (Fox, 1991), the best candidates appear to be amphiphilic molecules capable of spontaneously generating micelles and vesicles (Tanford, 1978; Luisi et al., 1999). There is evidence that lipid-like amphiphiles may have been imported or formed under prebiotic conditions (Gold, 1992; Ourisson and Nakatani, 1994; Deamer, 1997), although whether they existed in sufficient abundance is still open to debate (Cairns-Smith, 1982; Wachtershauser, 1990; Lahav, 1999). Lipid vesicles have also been shown to be capable of enhancing the rates at which precursors are converted into vesicle-forming amphiphiles (Bachmann et al., 1992). In some settings, this leads to an autocatalytic expansion of the molecular assemblies, a process resembling cell growth.

But how could such lipid assemblies carry and propagate information? The compositional genome model proposes that, early on, lipid-like compounds were extremely diverse, since they could be formed by combinatorial joining of diverse lipophytic tails and hydrophobic head groups (Segré et al., 2000b). An analogy to a combinatorial library of ligands in modern pharmacology (Cousins et al., 2000; Weber, 2000) suggests that millions of molecular configurations could occur. In the case of small assemblies (Figure 1), randomly seeded aggregates would manifest high compositional diversity, equivalent to a high information content (Segré and Lancet, 1999; Segré et al., 2000a). This is a sound departure point for early selection processes in a ‘Lipid World’ (Luisi et al., 1999; Segré et al., 2000b), utilizing the compositional genome replication mechanism.

Assemblies come to life

The graded autocatalysis replication domain (GARD) model (Segré et al., 1998a) provides a specific, quantitative description for the appearance and evolution of compositional genomes in a lipid world scenario. The model is based on a set of plausible chemical kinetic assumptions concerning mutually catalytic growth. Computer simulations demonstrate how amphiphilic assemblies manifest properties of homeostasis and compositional preservation (Segré et al., 1998a, 2000a). Monomer recruitment is governed by a set of kinetic differential equations that describe how mutually catalytic networks (Kauffman, 1986, 1993) could emerge from among the different compounds within the assembly (Segré et al., 1998b).

According to this model, each of the Nc compounds may serve to catalyze the recruitment or synthesis of some of the others, as governed by an NcXNc rate enhancement parameter matrix. The catalytic values are derived from published rate enhancement values (Fendler, 1982; Segré et al., 2000b) and from a statistical molecular recognition model (Lancet et al., 1993, 1994; Segré and Lancet, 1999). In this respect, lipid-like compounds, individually or in combination, may be considered as ‘lipozymes’ (Segré et al., 2000b) capable of generating primitive metabolic networks. At the same time, the involvement of lipid-like compounds at the very early steps of prebiotic evolution provides a means for a most natural appearance of compartmentalization (Figure 2).
A most rewarding facet of the GARD simulations for lipid assemblies are the new properties stemming from assembly splitting (Ségré et al., 2000a). The splitting assumption is plausible (Deamer, 1997; Norris and Raine, 1998) based on the known properties of soft matter, whereby through physical perturbation and external energy flux, larger assemblies might beget smaller ones. The computer simulations show the appearance of idiosyncratic, highly improbable compositional configurations ('composomes') (Ségré et al., 2000a) that are homeostatically stable for long periods of time. Eventually, these quasi-stationary states (Dyson, 1999) give way to new ones, as compositional mutations accumulate. This is analogous to the behavior of other biopolymer-free models (see also Wachtershauser, 1990).

It is noteworthy that, for identical kinetic parameters, different time courses arise due to minute changes in initial conditions, suggesting chaos-like properties. Within populations of composomes, groups with similar compositions often form, akin to quasi-species of replicating molecules (Eigen and Schuster, 1979; Kuppers, 1983), and these follow evolutionary dynamics of generation and extinction (Ségré et al., 1998a,b). The formation of self-organized amphiphilic assemblies occurs downhill energetically, and would reach a thermodynamic

---

**Fig. 1.** An assembly of amphiphilic molecules could spontaneously form according to two regimens governed by the following two parameters: \( N \), the number of molecules within an assembly (i.e. assembly size), and \( N_G \), the fixed, global number of different 'monomer' compounds attainable under the constraints of chemistry (i.e. repertoire size). For randomly formed large assemblies, fulfilling \( N >> N_G \) (top), every type of molecule is, on average, represented in multiple copies within each assembly, and therefore all assemblies are practically identical. By splitting, they form nearly identical replicas (+) but, because the assemblies manifest no diversity, evolution and selection are impossible. For small assemblies, where \( N << N_G \) (bottom), diversity is high, each molecular species is typically present in only one copy within an assembly and each assembly a different assortment of molecules. (For example, for an assembly containing \( N = 100 \) molecules and having a repertoire size of \( N_G = 1000 \), there are in excess of \( 10^{10} \) different assemblies possible, entering the realm of unlimited heredity (Szathmary and Maynard Smith, 1997). However, as previously pointed out (Morowitz, 1967; Szathmary and Maynard Smith, 1997), a split under the conditions \( N << N_G \) will invariably lead to two very different progeny, resulting in ineffective replication (bottom left, –). In order to reach an intermediate regimen in which both diversity and replication fidelity are achieved, it is necessary for the small assemblies to undergo a process of molecular repertoire restriction. The GARD model (see text) demonstrates that this is possible through the formation of organized, mutually catalytic networks (inset, A, thicker arrows). Importantly, repertoire restriction will not occur under conditions of equilibrium (inset, B) but a limited repertoire will be selected out of equilibrium (inset, C) (Ségré et al., 1998a,b). This process might happen only for a minute fraction within the large diversity of assemblies formed under the \( N << N_G \) conditions. Such assemblies will gradually reach a critical point, at which each molecule type will be present at two or more copies. Above this threshold, termed the 'Morowitz boundary' (Morowitz, 1967; Ségré and Lancet, 1999), effective replication by splitting would take place (bottom right, +), potentially leading to natural selection.
dead-end in the absence of a recirculation mechanism. In the GARD model, assembly splitting and disruption by turbulence or by thermal gradients keep the dissipative system far from equilibrium (see Nicolis and Prigogine, 1977; Morowitz, 1979). In other embodiments, involving the generation of molecular oligomers (Segré et al., 1998a; Imai et al., 1999), the generation of high-energy molecular precursors via chemical, thermal or light energy is shown to be equally crucial for coupling an energy source to an evolutionary process.

An important goal for future research will be to provide additional experimental bases for the compositional assemblies scenario. One could explore ways in which the assemblies would provide suitable microenvironments for diverse chemical reactions (Figure 2). The future will probably see large-scale and long time-span laboratory experiments, combined with microscopic chemical analysis of individual assemblies. At the same time, advances in in silico chemistry and biology over the next decade will make it possible to perform detailed computer simulations of molecular events within a large number of assemblies. Such a combination of ‘dry’ and ‘wet’ analyses could prove instrumental here, just as in other realms of biochemistry, e.g. cellular networks or protein folding.

Fig. 2. Comparison of two possible views for the path leading from a ‘primordial soup’ to a rudimentary protocellular structure (bottom). (A) The ‘biopolymer first’ scenario, according to which the emergence of self-replicating informational strings such as RNA and proteins are assumed to have had an independent origin from that of lipid encapsulation. (B) The ‘Lipid World’ scenario, which maintains that the roots of life could have been aggregates of spontaneously assembling lipid-like molecules endowed with capabilities for dynamic self-organization and compositional inheritance. More elaborate structures, including informational and catalytic biopolymers, might then have evolved gradually.
From assortments to alphabets

It will eventually be necessary to delineate mechanisms for the emergence of inheritance systems based on molecular ‘alphabets’ (Eigen and Schuster, 1979; Kuppers, 1983; Yockey, 1992). This should involve two distinct processes: the narrowing of monomer repertoire and the formation of polymeric strings (Blocher et al., 1999; Shapiro, 2000). In the traditional ‘bio-polymer first’ scenario (Figure 2A), both steps occur concomitantly, leading to the early emergence of self-replicating informational biopolymers such as RNA. The alternative view presented here (Figure 2B) invokes primordial compositional inheritance as a mechanism for self-reproduction, which could have started a process akin to Darwinian natural selection. This would provide a route for monomer selection prior to the appearance of polymers. Alphabet-based polymers might then become products of natural selection rather than prerequisite for it. Currently developed embodiments of the GARD model simulate the gradual accumulation of longer oligomers with rudimentary three dimensional structure. This leads to enhanced molecular recognition and catalysis, as seen in the realm of combinatorial biochemistry (Altreuter and Clark, 1999; Weber, 2000).

Conclusions

The chemistry of life as we know it today may have arisen from a set of chance events, analogous to those that led to the appearance of a specific body plan in organismal evolution. Future origin of life efforts should be devoted to charting general molecular evolution mechanisms, in analogy to deciphering the broad principles of developmental evolutionary biology. The models for early self-organization and reproduction delineated here could provide a conceptual framework for doing this, and might lead to quantitative predictions and suggestions for experimental tests. With the aid of disciplines including soft-matter physics, complex systems chemistry, astrobiology and nanotechnology, and the support of sophisticated computational analyses and simulations, this approach may help to address one of the most important challenges to modern science—understanding the way that life appeared on planet Earth.

Acknowledgements

We wish to thank Ora Kedem, David Deamer, Ayellet Falicovitz, Yoav Gilad, Gustavo Glusman and Roberto Ventrella for insightful discussions. Doron Lancet holds the Ralph and Lois Silver Chair in Human Genomics. Supported by the Crown Human Genome Center, The Israel Ministry of Science, Culture and Sports and the Krupp foundation.

References


Lancet, D., Kedem, O. and Pippel, Y. (1994) Emergence of order in small autocatalytic sets maintained far from equilibrium-application of receptor