

# Origins of Life: Concepts, Data, and Debates

*Managing Complexity Seems to be Essential for Studies on Chemical Evolution*

**D**ebates on the origin of life—or more precisely the terrestrial origin of life—as well as the origin of the universe are followed with great interest in almost all human societies. In the latter case, there exists a standard model, the big bang theory derived from an extrapolation of elementary particle physics to the birth of our universe. Nothing comparable is at hand for origin of life studies. Many different ideas are competing and none is available to provide a sufficiently plausible root to the first living organisms. It is even not clear what is meant by ‘life’ and possible definitions are heavily debated. What will be pointed out in this essay concerns two issues: First, no definition of life is needed for work on early or chemical evolution, and second, new approaches are required, and borrowing ideas from handling complex systems might be useful.

A list of criteria to be used in the classification of prelife and life could, for example, contain:

- i. multiplication and inheritance,
- ii. variation through imperfect reproduction and recombination,
- iii. metabolism for the production of molecular building blocks,
- iv. individualization through enclosure in compartments,
- v. homeostasis and autopoiesis,
- vi. organized cell division (bacterial cell division or mitosis),
- vii. sexual reproduction and reductive division (meiosis), and
- viii. cell differentiation in germ line and soma.

To give examples, viroids fulfill only criteria (i) and (ii), viruses (i), (ii), and (iv), and bacteria all criteria from (i) to (vi). Even artifacts can be readily classified: Computer viruses meet criterion (i) and computer worms fulfill (i) and (iv). It is interesting to note that the current computer attacking artifacts do not meet criterion (ii): their evolution is fully in the hands of the “hackers” creating them.

Biological evolution has always two aspects that are both required for understanding present day life: (i) the historical dimension recording how current living forms were derived from earlier species, and (ii) the mechanistic dimension explaining how the evolutionary process works. The history of evolution is tantamount to the fossil record and its interpretation. As far as origin of life questions

**PETER SCHUSTER**

*Peter Schuster is at the Institut für Theoretische Chemie der Universität Wien, A-1090 Wien, Austria (e-mail: pks@tbi.univie.ac.at)*

are concerned, the fossil record is of practically no use since the oldest fossils that are believed to be of biogenic origin are about 3.5-billion-year old and represent remnants from organisms resembling “stromatolites” formed by present day cyanobacteria [1]. Although some doubts were raised whether or not this interpretation is correct [2], the evidence for a biological origin of these archaean stromatolites is very strong [3]. For the arguments used here the age of the oldest witnesses of life makes no difference, because at our current state of knowledge nothing is available in the fossil record that could provide a hint on the pathways of chemical evolution. Today, the historic route to life is entirely in the realm of speculation and, presumably, will stay there forever.

Works on reaction mechanisms that are considered to be important for origin of life questions can never be falsified or confirmed, but they can be classified as implausible or plausible. Ever since the early works of Alexander Oparin in 1920 on “coacervates”,<sup>1</sup> scientists interested in the primordial world have constructed an impressive number of scenarios [4]. Almost all present day researchers on origin-of-life problems will agree that a primitive cell requires three components, a molecular basis for genetics being RNA or DNA, metabolism, and compartmental structure with a division mechanism. The debate focuses primarily on the (temporal) sequence in which

<sup>1</sup>*Coacervates are small droplets formed from fatty acids, proteins, or other amphiphilic polymers exposed to aqueous media. They may grow and split into smaller droplets and are considered as potential candidates for precursors of organized vesicles or primitive cells. “Amphiphilic” means that the molecules consist of two parts: the hydrophilic part, which dissolves readily in water and the hydrophobic part that has very little or no solubility in water-like oil.*

the three components were established. With some simplification most of these scenarios fall into three major classes: (i) genetics first models, (ii) metabolism first models, and (iii) compartments first models. All three scenarios suffer from major unexplained details and rather few aspects concerning the transition from prebiotic chemistry to early life are not debatable. To give one example modern life is based on replication of genetic information that can be traced into the past by the means of phylogenetic analysis. No matter what the scenarios on the prebiotic Earth were, replication and phylogeny began 1 day and therefore sentences like, “A replicator was not involved in the origin of life” [5] are dispensable, misleading, and express nothing but the distaste of the author for a definition of life that includes genetics. One major difference between the various scenarios is concerned with the question, whether the path towards early life was autotrophic in the sense that a prebiotic metabolism provided the required materials or heterotrophic with the implicit assumption that the required materials were in the environment as a result of other processes, and metabolism in the sense of modern biochemistry was developed later by early forms of protolife to become independent of environmental coincidences.

Genetics first models agree that an RNA world [6] preceded our present day DNA-RNA-protein world, but despite some recent progress [7], it is not known how the first RNA molecules originated under prebiotic conditions. On the other hand, enzyme-free cross-catalytic synthesis of RNA-based enzymes, called ribozymes, was recently shown to result in exponential growth and selection [8], and there is little doubt that Darwinian evolution could have started in an RNA world without protein catalysis. A very rich repertoire of diverse catalytic functions of ribozymes is known at present [9]. RNA molecules are especially effective

in processing nucleic acid and protein molecules.

Metabolism first models are very attractive because they might give an answer how prebiotic chemistry was canalized to provide a suitable reservoir for the abiotic synthesis of biomolecules [10]. As metabolism within a compartment, when driven by an external source of energy and inorganic materials, could provide all necessary building blocks for protocell proliferation, such metabolisms would be autotrophic. Most of the current metabolism first models are based on the suggestive argument that catalytic cycles selected a subset from the overwhelmingly diverse repertoire of organic molecules that could arise under prebiotic conditions [11–13]. Unfortunately, almost all proposed models suffer from the lack of experimental implementation and the role of larger cycles at the origin of life appears to be rather implausible or at least questionable [14, 15]. In particular, catalytic cycles with many components like the tricarboxylic acid cycle and its inversion require highly specific catalysts that seem to be inaccessible without protein enzymes. High yield and high specificity is indeed indispensable already for cycles with moderate numbers of intermediates: For the purpose of illustration, we suppose a cycle of 10 reactions each of which occurring with a respectable yield of 80%. One full rotation of the cycle provides a yield of only 11%, 89% are byproducts. Without minimization of the loss of material through side reactions larger cycles cannot work. One well-studied experimental example of an autocatalytic reaction network is the formose reaction [16], which converts formaldehyde into carbohydrates. It operates with high efficiency but gives rise to a highly diverse mixture of different compounds. Interestingly, catalysis by borate was found to lead to some preference for ribose [17]. The design of appropriate candidates for plausible prebiotic metabolisms encounters two

major difficulties: (i) The weakness of our current theories to predict the properties of dynamical networks, and (ii) systematic search on the experimental properties of small and medium size (auto)catalytic reaction networks. The first problem is shared with the theory of complex systems.

Compartments first models are based on the fact that amphiphilic<sup>1</sup> molecules are available under prebiotic conditions. They form a great variety of aggregates in aqueous solutions. Such aggregates range from micelles with hydrophobic material inside and hydrophilic material on the interface to water to double layer vesicles with water inside and outside separated by a membrane formed by the hydrophobic parts of the amphiphilic molecules. Vesicles are particularly interesting because their membranes have basic features in common with cell membranes. Simpler than vesicles are aggregates of amphiphilic materials and a theoretical concept making a point for self-enhancing formation and evolution of micelle-like aggregates called “composomes” has been derived [18–20]. A composome theory, the Graded Autocatalysis Replication Domain (GARD) model, has been worked out [19] and attempts to conceive an experimentally addressable, real GARD model are under way. Recent analysis of composome theory revealed, however, a serious lack of evolvability in the GARD model [21]. The model systems do not create new favorable aggregates but iterate between forms that were—although somewhat hidden—already present in the initial assumptions on the composome clusters.

There is a related fundamental obstacle for composome evolution that is shared by autocatalytic sets of proteins, which engage in mutual or cross catalytic reproduction [22]. Mutation in both systems is introduced by an accidental error in composome assembly or in protein catalysts for protein synthesis. If, fortuitously, the mutation is advantageous it will give rise to a more efficient composome or autocatalytic

cycle, and this entity will produce more. In itself, however, it is not autocatalytic and cannot enhance its own production. Moreover, when the error was a rare event it is not reproducible, because the system has no inheritance in the sense of genetic information that can be copied on demand [23]. Digital information and information storage as it occurs with nucleic acids is a “*conditio sine qua non*” for evolution and composomes as well as autocatalytic cycles are lacking it. In case the mutation is a very frequent event, evolution is impossible either, because the mutant is just one more component of the equilibrium mixture.

Eight years ago a program to synthesize artificial cells or protocells that capture the essential properties of life has been proposed [24]. The basic idea was to encapsulate genetic material in vesicles that allow for replication and correlated vesicle division—according to the initially given criteria such protocells would fulfill (i), (ii), and (iv). Artificial cells that allow for enzyme-catalyzed replication and vesicle division were successfully assembled already earlier [25]. It is worth mentioning that more complicated molecular machineries including translation of RNAs into protein were also proven to work in liposomes [26]. Recently, it was demonstrated that vesicles built from fatty acids allow for the uptake of charged activated nucleotides [27] from the surrounding aqueous solution. These nucleotides were used in enzyme-free template-directed synthesis of genetic polymers that was followed by vesicle division. These investigations provide a “proof of concept” for the possibility of a heterotrophic origin of the first cells. A great variety of protocells that meet other conditions of early life, for example intracellular metabolism (iii), have been produced, and for details we refer to an excellent collection of recent reviews [28].

Information on the evolution of metabolism after the RNA world is eas-

ier accessible. A recent approach towards early metabolism is based on comparative genomics and protein structure analysis. Phylogenomics of protein architecture suggests an evolutionary sequence for the appearance of the various modules in the networks of present day metabolism [29, 30]. According to these investigations the first enzymatic takeover from prebiotic chemistry was related to the synthesis of nucleotides for the RNA world. Further takeovers concerned the metabolic pathways leading to amino acids, carbohydrates, and eventually lipids. Afterwards, fast diversification into proteins of the three superkingdoms—eubacteria, archaea, and eukarya—occurred. Together with the results on template-directed replication in protocell-like vesicles, the data on modern metabolism speak for a heterotrophic origin of life: Building blocks are assumed to have been available on the prebiotic Earth and a protocell could import them. It seems to be easier to be fussy in the uptake of the right compounds than to canalize prebiotic chemistry to produce the required materials. Nevertheless, enrichment of a primordial soup in the right molecules certainly was required as well.

Despite rather pessimistic views uttered by the opponents of a natural origin of life an impressive collection of data is available now. They all speak for a sequence of prebiotic events and processes leading from networks of dynamically related small molecules and amphiphiles to biological macromolecules, compartments, and eventually protocells. Not a single one of the suggested avenues to life seems to be plausible but taking them all together in a concerted view could provide the solution. Within vesicles RNA world scenarios, assisted by Darwinian evolution, initiated further development towards modern biochemistry based upon DNA, RNA, and protein, as well as the familiar protein catalyzed metabolism. At present, this concept is a patchwork with a number of still miss-

ing pieces to complete the puzzle. Apart from more experimental work, a new and comprehensive theory of chemical systems is required that allows for straightforward analysis of network dynamics, a need that is shared with complex systems theory and its applications to problems in other fields.

## REFERENCES

- Schopf, J.W.; Packer, B.M. Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. *Science* 1987, 237, 70–73.
- Brasier, M.; Green, O.; Lindsay, J.; Steele, A. Earth Oldest (~3.5 Ga) fossils and the “early eden hypothesis”: Questioning the evidence. *Orig Life Evol Biosph* 2004, 34, 257–269.
- Schopf, J.W. Fossil evidence of Archean life. *Philos Trans R Soc (London) B* 2006, 361, 869–885.
- Hazen, G.M. *Genesis: The scientific quest for Life's origins*; Joseph Henry Press: Washington, DC, 2005.
- Shapiro, R. A replicator was not involved in the origin of life. *IUBMB Life* 2000, 49, 173–176.
- Gestland, R.F.; Cech, T.R.; Atkins, J.F., Eds. *The RNA World. The Nature of Modern RNA Suggests a Prebiotic RNA World*, 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2006.
- Powner, M.W.; Gerland, B.; Sutherland, J.D. Synthesis of activated pyridine ribonucleotides in prebiotically plausible conditions. *Nature* 2009, 459, 239–242.
- Lincoln, T.A.; Joyce, G.F. Self-sustained replication of an RNA enzyme. *Science* 2009, 323, 1229–1232.
- Lilly, D.M.; Eckstein, E., Eds. *Ribozymes and RNA Catalysis*; Royal Society of Chemistry Publishing: Cambridge, UK, 2007.
- Eschenmoser, A. The search for the chemistry of life's origin. *Tetrahedron* 2007, 63, 12821–12843.
- Morowitz, H.J.; Kostelnik, J.D.; Yang, J.; Cody, G.D. The origin of intermediary metabolism. *Proc Natl Acad Sci USA* 2000, 97, 7704–7704.
- Smith, E.; Morowitz, H.J. Universality in intermediary metabolism. *Proc Natl Acad Sci USA* 2004, 101, 13168–13173.
- Schuster, P. Taming combinatorial explosion. *Proc Natl Acad Sci USA* 2000, 97, 7678–7680.
- Orgel, L.E. Self-organizing biochemical cycles. *Proc Natl Acad Sci USA* 2000, 97, 12503–12507.
- Orgel, L.E. The implausibility of metabolic cycles on the prebiotic Earth. *PLoS Biol* 2008, 6, e18.
- Breslow, R. On the mechanism of the formose reaction. *Tetrahedron Letters* 1959, 1, 22–26.
- Ricardo, A.; Carrigan, M.A.; Olcott, A.N.; Benner, S.A. Borate minerals stabilize ribose. *Science* 2004, 303, 196.
- Segrè, D.; Lancet, D.; Kedem, O.; Pilpel, Y. Graded autocatalysis replication domain (GARD): Kinetic analysis of self-replication in mutually catalytic sets. *Orig Life Evol Biosph* 1998, 28, 501–514.
- Segrè, D.; Ben-Eli, D.; Lancet, D. Compositional genomes: Prebiotic information transfer in mutually catalytic noncovalent assemblies. *Proc Natl Acad Sci USA* 2000, 97, 4112–4117.
- Hunding, A.; Kepes, E.; Lancet, D.; Minsky, A.; Norris, V.; Raine, D.; Sriram, K.; Root-Bernstein, R. Compositional complementarity and prebiotic ecology in the origin of life. *BioEssays* 2006, 28, 399–412.
- Vasas, W.; Szathmáry, E.; Santos, M. Lack of evolvability in self-sustaining autocatalytic networks: A constraint on metabolism-first path to the origin of life. *Proc Natl Acad Sci USA*, in press.
- Kauffman, S.A. Autocatalytic sets of proteins. *J Theor Biol* 1986, 119, 1–24.
- Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 1971, 58, 465–523.
- Szostak, J.W.; Bartel, D.P.; Luisi, P.L. Synthesizing life. *Nature* 2001, 409, 387–390.
- Oberholzer, T.; Wick, R.; Luisi, P.L.; Biebricher, C.K. Enzymatic RNA replication in self-reproducing vesicles: An approach to a minimal cell. *Biochem Biophys Res Comm* 1995, 207, 250–257.
- Kita, H.; Matsuura, T.; Sunami, T.; Hosoda, K.; Ichihashi, N.; Tsukada, K.; Urabe, I.; Yomo, T. Replication of genetic information with self-encoded replicase in liposomes. *ChemBioChem* 2008, 9, 2403–2410.
- Mansy, S.S.; Schrum, J.P.; Krishnamurthy, M.; Tobé, S.; Treco, D.A.; Szostak, J.W. Template-directed synthesis of a genetic polymer in a model protocell. *Nature* 2009, 454, 122–126.
- Rasmussen, S.; Bedau, M.A.; Chen, L.; Deamer, D.; Krakauer, D.C.; Packard, N.H.; Stadler, P.F., Eds. *Protocells. Bridging nonliving and living matter*; The MIT Press: Cambridge, MA, 2009.
- Caetano-Anollés, G.; Kim, H.S.; Mittenthal, J.E. The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc Natl Acad Sci USA* 2007, 104, 9358–9363.
- Caetano-Anollés, G.; Yafremava, L.S.; Gee, H.; Caetano-Anollés, D.; Kim, H.S.; Mittenthal, J.E. The origin of modern metabolism. *Int J Biochem Cell Biol* 2009, 41, 285–297.