Chapter 16

EMERGENCE AND SELECTION OF BIOMODULES: STEPS IN THE ASSEMBLY OF A PROTOCELL

Susanna C. Manrubia and Carlos Briones

Centro de Astrobiología (INTA-CSIC)
Ctra. de Ajalvir km. 4, 28850 Torrejón de Ardoz, Madrid, Spain
scmanrubia@cab.inta-csic.es

Life was probably ubiquitous on Earth some 3,500 million years ago. In a period that could have been as short as 100 million years, cells very similar in structure and metabolism to extant ones had developed from abiotic matter. The pathway from inorganic chemistry to the first self-replicating molecule and the first functional metabolism, both enclosed in a membrane-based compartment, is marked by major difficulties such as a limited availability of simple biomodules, a too slow reaction rate between molecules, the inefficient formation of short homochiral polymers, an unfaithful template copy, or the fragility of the initial self-sustained catalytic cycles. Miller and Urey's experiment in 1953 moved the problem of the origin of life to the experimental sciences. Since then, significant advances have been accomplished, among them the identification of several RNA molecules which were able to catalyze essential biochemical reactions or the design of protocells able to grow and divide. Experimental achievements are intimately linked to technological advances and to the development of increasingly more realistic theories and models addressing the different stages involved in the origin and evolution of complex chemistry and early life.

Contents

1. Preamble

This chapter contains a non-exhaustive revision of our current knowledge on prebiotic chemistry as well as on the combination of genetic molecules, protometabolic cycles and membrane-based compartments in the pathway from simple building blocks to the first protocells. The subject is so vast that, from the onset, we renounced any attempt to be historically complete or to cover all of the many chemical aspects involved. Instead, we present the basic scenario, the motivation to pursue some key experiments, and some of the conceptual problems that research on prebiotic chemistry and life's origins has uncovered.

The development of a physical environment which is able to sustain complex chemistry is prior to the emergence of life as we know it: the early Earth had to be transformed for life to emerge. We discuss some of the difficulties that, in a hypothetical pathway from inorganic to organic chemistry, should be overcome to generate self-replicating and self-sustained chemical systems. As an illustration of the many efforts made over the last 60 years, we will highlight recent experimental developments and new scenarios that are broadening the way we think about prebiotic chemistry and chemical organization. Admittedly, it is unlikely that we will ever discover how life actually originated. We would be satisfied to find out, maybe in a not so further future, a single continuous pathway leading to the emergence of a protocell of the simplest kind.

2. Introduction

2.1. *The early Earth*

Building a habitable planet takes a long time. Our Earth-based knowledge suggests that it takes significantly longer than it is later needed for

the planet to become inhabited by microorganisms. The transition from inorganic matter to life on Earth could have taken place quite rapidly in a geological scale once our planet cooled down sufficiently for liquid water to be stable on its surface.

Four and a half "giga-annum" (Ga) or billion years ago, the Earth was a hot conglomerate of melted rocks, gas and dust. Before life could emerge and develop, the physical and geological features of our planet had to be severely modified. The crust of the Earth differentiated from the core relatively early, some 4.4 Ga ago. A magnetic field developed afterwards, and, in a multi-stage process that witnessed the addition of different volatiles, a primitive atmosphere composed mainly of nitrogen, carbon dioxide and water vapor, eventually set in. The apperance of oceans was strongly dependent on the delivery of water from extra-terrestrial sources such as comets and meteorites, and on the presence of an atmosphere thick enough to prevent excessive escape of water vapor. There is evidence that proto-oceans might have been present as early as the first crust appeared. At that time, the atmospheric pressure could have reached 273 bars, maintaining oceans at a temperature higher than 230°C.[1] The terrestrial crust was strongly rebuilt for about 500 million years. In fact, although the oldest terrestrial material is a zircon mineral dated 4.4 Ga ago, the oldest rock we know is 4.1 Ga ago. The last major event probably preventing (or hindering) the appearance of life was the Late Heavy Bombardment, which took place between 3.95 and 3.8 Ga ago.

Therefore, between the formation of Earth (4.55 Ga ago) and the generation of a habitable surface, some 700 million years elapsed. The precise timing of the events that took place before life appeared is impossible to tell. Indeed, it is likely that life originated (and went extinct) several times during that period, being current organisms the outcome of the only successful trial. There is some consensus in that a prebiotic chemistry stage could have started some 4.4 to 4.2 billion years ago. At that point, the terrestrial environment was ready for chemical reactions able to produce significant yields of relatively complex molecules that, we believe, were subsequently used in the emergence of self-sustained, self-replicating systems. Indirect evidence of life is found some 3.8 Ga ago, when a fractionation of carbon isotopes compatible with biological activity is observed. However, the first incontrovertible fossils are dated to 3.5 Ga ago.[56]

2.2. *Approaches to early life*

Looking at the analogies among extant organisms, Darwin inferred, a century and a half ago, the likely existence of a common ancestor from which we all, from tiny bugs to primates, derive through modification and

selection. That morphological analogy turned into genetic relatedness once DNA was identified as the material carrying the inheritable information. The last universal common ancestor (LUCA) of the three phylogenetic domains (Archaea, Bacteria and Eukarya) lived between 3.8 and 3.5 Ga ago, as the genes of ribosomal RNA and many other of our shared genes demonstrate.[57] However, chemistry was probably tinkering with prebiotic molecules as far as 4.4 Ga ago. Little to none evidence of the steps accomplished before LUCA arrived is left. The fossil record does not preserve molecular aggregates or evolving entities smaller than cells. Also, geochemical biomarkers do not inform about past life longer than some 2.5 Ga ago. In turn, relevant clues can be derived from the current biochemical functions: some of the so-called 'molecular fossils', such as ribozymes, may tell us about the features of molecular evolution before LUCA. But the fact is that we are trying to reconstruct the pathway from inorganic chemistry to the first protocells mostly in the absence of empirical evidence.

Approaches to study prebiotic chemistry and the path to life either try to find out how the building blocks of biochemistry could have been generated from inorganic compounds (starting 4.4–4.2 Ga ago) or look at extant cells and reduce them in order to disentangle how a minimal cell might have looked like (going back to 3.8 Ga ago, at best). There is a large gap in time and in chemical complexity between a bottom-up approach (the first case) and top-down research (the latter). And both ways of looking at the origins of life are strongly conditioned by life as we know it.

In 1952 Stanley Miller performed an experiment in Harold Urey's laboratory that inaugurated the empirical age of prebiotic chemistry. They demonstrated that several proteinogenic amino acids and other essential biomolecules could be abiotically produced. It seemed that the path to synthesizing life in the lab was paved, that the big question of life's origin would be answered shortly after. Unfortunately, there were many major issues to be sorted out, among them the separation of racemic mixtures of products into homochiral subsets, the synthesis of some compounds essential in modern cells but of low stability (as certain sugars), the concatenation of monomers (e.g. amino acids and nucleotides) into polymers (proteins and nucleic acids), the emergence of a replication machinery or the generation of appropriate cell-like compartments, not to mention catalytic networks, required to establish a primitive metabolism. More than half a century after Miller–Urey's experiment, we are still far from completely solving the problem. But we certainly have moved forward.

Regarding the top-down approach, it is evident that modern cells are extremely complex molecular factories able to perform a huge number of

different functions. The number of genes that a hypothetical minimal cell requires to complete its basic functions is estimated between 50 and 400.[2,58] Lower estimates only include synthesis of DNA, RNA and proteins, and assume that membrane formation and cell division could occur thanks to the action of physico-chemical processes not coded in the cellular genome. Further, this line of research often seeks to understand the performance of small gene circuits and how they can be integrated into functional primitive metabolisms, in a systems biology-like approach. This view is closer to the design of an artificial cell than to the characterization of a true minimal cell. Upper estimates (from 300 to 400 genes) begin with extant cells and try to reduce the number of genes by dispensing apparently non-essential functions. This reduction is qualitative and retrieves minimal cells far too complex to be understood from first constructive principles.

3. Synthesis and Accumulation of Biomodules

Before the appearance of life, our planet was a very complex chemical reactor. Plate tectonics appeared early in Earth's history, so one could move from oceans to the summit of mountains through often tempestuous coasts and experience broad variations in temperature. The action of tides created cycles of wetting and desiccation, igneous rocks were present from the very beginning, and sedimentary rocks were forming at least 4.1 Ga ago. The volcanic activity at the bottom of oceans formed submarine vents that created strong temperature and solvent concentration gradients. Such a broad spectrum of environments is impossible to recreate in the laboratory, not to mention the hundreds of millions of years during which the huge terrestrial laboratory was playing with chemistry. The logical inference is that a bunch of different prebiotic experiments were simultaneously taking place, thus transforming the available substrates into complex mixtures of products. Many of those experiments probably were dead ends, others maybe led to self-sustained, replicating systems that were later outcompeted by more efficient solutions. It cannot be discarded that different pieces of what eventually was the first protocell were generated in different environments and later combined. Current attempts to recreate the origins of life should therefore consider this complex (and more realistic) scenario and follow a "systems chemistry"-based approach.[59] In any case, some 300 million years before the first protocells, stable and active enough biomodules had to be produced, physico-chemically separated from accompanying compounds, and accumulated in a reusable form.

3.1. *Life's origin in the lab: Miller and Urey's experiment*

In the decade of the 1920's, Alexander Oparin and John B. S. Haldane had hypothesized that conditions on the primitive Earth should have permitted the synthesis of organic compounds from inorganic precursors. Three decades later, Miller and Urey decided to test this hypothesis. At their time, the early terrestrial atmosphere was believed to be highly reductive, so they took a mixture of water, methane, ammonia, and hydrogen to recreate it. The activation energy required for chemical reactions to occur was delivered by two electrodes that mimicked lightning through the atmosphere. The reactions took place inside a sealed flask containing the four gases, and the condensed products were collected in a second flask containing liquid water. After a few days, a significant fraction of the carbon in the system was forming organic compounds, with about 2% of it contained in amino acids, glycine being the most abundant.[3] Miller and Urey's spark discharge experiment has been repeated and re-examined several times to confirm and enlarge the repertoire of organic compounds produced: more than 20 different amino acids and a number of other biomolecules have been abiotically synthesized.[4]

Afterwards, several alternative environments have been tested in the laboratory. Subsequent evidence indicated that early Earth's atmosphere was not as reducing as previously believed, and due to volcanic activity it likely contained significant amounts of nitrogen, carbon dioxide, hydrogen sulfide, and sulfur dioxide. Under these conditions different repertoires of molecules, including amino acids and nitrites, are produced. One of the effects of the latter is, unfortunately, to rapidly destroy amino acids. However, the addition of iron and carbonate minerals, likely abundant in the early Earth, reverses that effect, and proteinogenic amino acids are again obtained in significant yields.[5]

Other experiments have explored atmospheres abundant in carbon monoxide and molecular hydrogen, the effect of UV light as a source of energy, and the role that low temperature could have played with regard to the stability of the reaction products. The production of amino acids and other biomodules in significant amounts seems to be unavoidable in any mixture of simple (atmospherically plausible) volatiles exposed to an energy source. Most of the natural amino acids, purines, pyrimidines, and sugars appeared in different variants of the original Miller–Urey experiment.

The first step towards the abiotic generation of life has been firmly advanced. Still, the mixtures so produced are highly heterogeneous and racemic. As we will discuss in Sec 3.3, the selection of the chirality of the biomodules is prior to the polymerization of biological macromolecules.

3.2. *Sugars and nucleotides*

Some molecules essential in extant biochemistry, sugars and nucleotides, are not that easily obtainable by abiotic means. Monosaccharides, the simplest sugar molecules, are relevant biomodules in living systems, playing key roles in metabolism and rendering structural or energy rich polysaccharides upon polymerization. Among them, ribose is an important component of many coenzymes and constitutes, together with phosphate, the molecular backbone of RNA. In turn, DNA is characterized by its deoxyribose-phosphate backbone. Therefore, the prebiotic synthesis of ribose (and, to a lesser extent, deoxyribose) was soon explored due to their key role in the polymeric genetic macromolecules. Formaldehyde was found to oligomerize in the presence of mineral catalysts to form sugars (in the classical formose reaction, discovered by Butlerow in 1861), although complex, tar-like mixtures of tetroses, pentoses and hexoses are obtained, being ribose a relatively minor product.[60] The presence of borate minerals stabilizes ribose in the mixture of sugars, what suggests a plausible mechanism for the accumulation of the precursor of ribonucleotides.[61]

The synthesis of ribonucleotides had been pursued for over 40 years under the assumption that they should assemble from their three molecular components: ribose, a nucleobase and phosphate. In this scenario, ribose and nucleobases would have been produced independently and then combined. The first synthesis of a purine nucleobase (adenine) was achieved in 1960 through the polymerization of HCN.[62] Since then, the formation of purines (and, less efficiently, pyrimidines) has been achieved under different conditions including eutectic phases, ice matrices and drying/wetting cycles. Mineral and metal surfaces also enhance these processes by concentrating reagents and preventing products from degradation.[63] Nevertheless, two very unfavorable reactions are required in order to produce a ribonucleotide: the formation of a glicosidic bond between ribose and the nucleobase, and the phosphorilation of the resulting nucleoside. In particular, no way of joining ribose and canonical pyrimidines has been ever found.

Recently, an alternative solution to the pyrimidine+ribose equation leading to the corresponding ribonucleosides has been found. The way out of that conundrum required to escape the old assumption of independent synthesis of nucleobases and ribose and, following a novel "systems chemistry" approach, to look for a path with a common precursor.[11] This likely chemical pathway leading to ribonucleosides starts, as many other prebiotic reactions, with very simple building blocks including glycolaldehyde and cyanamide. Interestingly, when inorganic phosphate is added to the mixture, most of the unwanted reactions are eliminated

and the key intermediate, 2-aminooxazole, is efficiently synthesized. This compound, which contributes both to the sugar and the nucleobase motifs of the ribonucleotide, is volatile enough to be purified by sublimation–condensation cycles. These cycles are by no means complex: day-night variations could suffice to accumulate 2-aminooxazole, a first step in RNA synthesis. The further phosphorilation of the ribonucleoside is facilitated by the presence of urea, which comes from the phosphate-catalyzed hydrolysis of cyanamide. This leads to the final production of pyrimidine ribonucleosides.

3.3. *Chirality*

Abiotic chemical reactions where chiral molecules are produced usually yield levorotatory or "left-handed" (L) and dextrorotatory or "right-handed" (D) enantiomers with equal probability. But long polymers, believed to be at the basis of most genetic systems, are only possible if formed exclusively by one of the two enantiomeric types. Life uses L-amino acids and D-sugars: the mechanisms that broke the symmetry of racemic mixtures and yielded enantiomerically enriched products have puzzled researchers for decades.[64] As with other accomplishments in prebiotic chemistry, we cannot be certain that the solutions found in the laboratory are those that Nature used. But some plausible, and not necessarily complex, scenarios have been devised: Mixtures with an enantiomeric excess over 99% can be achieved by simply stirring a racemic solution.[6]

At the root of the abiotic generation of homochirality lie two processes that were probably essential also at other stages: the (autocatalytic) chemical selection of molecules with a certain property (handedness in this case) and their accumulation for possible later use. The theoretical prediction that mixtures with a large enantiomeric excess should result from a process where each enantiomer would catalyse its own production was contemporary to Miller–Urey's experiment,[7] though its experimental demonstration had to wait for half a century.[8] In those experiments, the system could not achieve a 100% enantiomeric excess. A theoretical way out could be provided by recycling the less abundant enantiomer to the most abundant type, which would accumulate in crystal form until chiral purity is achieved.[9] Shortly after, it was experimentally demonstrated with initially racemic mixtures of sodium chlorate that this process is possible, efficient, and leads indeed to the accumulation of chiraly pure compounds.[10] In any case, a growing number of alternative phenomena have been theoretically postulated or experimentally tested to have generated chiral biomodules (and their polymers) from non-chiral matter.

3.4. *Prebiotic polymerization of biomodules*

The polymerization of biomodules (mainly, amino acids and nucleotides) into polymeric macromolecules (polypeptides and nucleic acids) in the absence of enzymes had to deal with the thermodynamically uphill process of water removal, required for condensation reactions. To overcome this limitation, different prebiotic scenarios have been considered, including hydration–dehydration cycles and melting processes, as well as the presence of heterogeneous systems containing either mineral surfaces or lipid domains. Also, these systems should provide a favorable environment for the stabilization of the polymer against hydrolysis once the biomodules have been condensed.

The synthesis of polypeptides from amino acids required the conversion of peptide bond formation into a thermodynamically favorable process. Different mechanisms have been proposed, and some have been experimentally tested. The most successful experimental settings involve fluctuating heating cycles in the presence of mineral surfaces (e.g. silica, alumina and the montmorillonite clay),[65] the presence of small organic activating molecules (such as imidazol or carbodiimides) in combination with mineral surfaces at low temperature,[66] and wetting/drying cycles in the presence of concentrated NaCl solutions and Cu^{2+} as a metal catalyst.[67] Additionally, the use of activated amino acids instead of their natural forms enhances the polymerization rate and the polypeptide length.[68]

Regarding the non-templated polymerization of ribonucleotides, the activation of the phosphate with different leaving groups (mainly nitrogen-containing heterocycles, such as imidazole, pyridine or purine derivatives) has been assayed. The longest RNA polymers (up to 50-mers) have been obtained using imidazole- and 1-methyladenine-activated ribonucleotides — thanks to the concentration and catalytic properties of montmorillonite interlayers.[69] In turn, non-activated ribonucleotides can polymerize up to 25- to 100-mers at high temperature, in a dehydration/rehydration system containing fluid lipid matrices composed of amphiphilic molecules.[70]

Therefore, a plausible scenario might have involved the dynamic interaction of different biomodules with the montmorillonite clay: the phyllosilicate surfaces or interlayers could have promoted, in contact with the bulk aqueous medium, the synthesis of nucleobases, the polymerization of (activated) ribonucleotides and that of amino acids. The cooperation of this system with the micro-environments provided by amphiphilic-based vesicles is also favored by the experimental evidence. This fact highlights the relevance of heterogeneous catalysis in the origins of life.

3.5. *Chemical selection and compartmentalization*

The rate of chemical reactions occurring among the different small biomodules generated through abiotic reactions requires that they were present in a sufficiently high concentration and, preferably, in the vicinity of other chemical species that could act as catalysts. Montmorillonite and other mineral surfaces could have been essential in the production of polymers: they may act by selecting and, in a sense, compartmentalizing the universe of possible reactions among the building blocks of macromolecules.

The many different environments simultaneously present on the early Earth surely provided opportunities for synthesis, differential selection, compartmentalization and, occasionally, catalysis of simple biomodules. An interesting setting occurs at the bottom of oceans, in volcanically active regions where hydrothermal vents form. High pressures maintain water in liquid state occasionally above 400°C. Surrounding waters rapidly cool down in a gradient that sustains a rich chemical activity[12] and a remarkable biodiversity. Since the discovery of black smokers in the late '70s of the past century, the interest in those submarine formations has steadily increased, to the point that a hydrothermal origin of life has been proposed.[13] The abundance of reduced organic compounds, the complex biogeochemistry of those areas, a continuous and concentrated source of energy, the presence of active hydrothermal systems 4.2 Ga ago, or the many lithotrophic microorganisms described these environments as attractive and challenging regions of study in relation to life's origin.[14]

Hydrothermal vents have additional advantages to foster the selection and compartmentalization of complex chemical compounds. Thermal gradients inside vent channels promote the separation and differential accumulation of biomodules, both through passive thermal diffusion and enhanced by convection. Interestingly, convection also creates thermal cycles that could be vital to promote complex biochemical reactions, in the same way that polymerase chain reaction (PCR) techniques lead to DNA amplification in current laboratories.[15] Further, rocks of volcanic origin are highly porous: the abundance of natural compartments in probably stable physico-chemical conditions could have fuelled further reactions among the chemical species accumulated in such micro-environments. Nevertheless, hypotheses favoring a "hot origin" of life have to face important problems related to the reduced stability of most biomolecules at high temperatures.[71]

In other environments, still, amphiphilic molecules (e.g. single-chained fatty acids) plausibly synthesized in certain prebiotic scenarios[72] may have spontaneously self-assembled into membrane-based vesicles (later called liposomes when they were formed by complex lipids) which in turn could have compartmentalized the first replicating molecules and/or

protometabolic cycles. Vesicles are relatively stable at a wide range of sizes, being able to grow by the slow addition of fatty acids in the form of micelles. Nevertheless, when a maximum volume-to-surface ratio is reached (or if physico-chemical perturbations affect the stability of the system) the vesicle may divide into two daughter vesicles[73] and distribute its internal content between the offspring. This process would have constituted a primordial, very simplified and unregulated version of proto-cellular division.

Most scientists would accept that a rich repertoire of molecular compounds and membrane-forming amphiphiles can be produced abiotically. However, the subsequent steps, which are thought to be the advent of template replication, the emergence of autocatalytic reaction networks and the coupling between proto-genome replication and compartment reproduction are major transitions in evolution[16] that demand a conceptual shift.

4. Template Replication and the RNA World

In the context of the origin of life, replication can be defined as any reliable copying process of a polymeric template whose outcome is a new molecule which preserves the specific sequence of the template.[88] The first true non-enzymatic self-replicating experimental system used a palindromic DNA hexamer which assisted the ligation of two DNA trimers — each of them complementary to one half of the template.[23] Another relevant example of self-replication used 15-mer and 17-mer oligopeptides that covalently bind to each other through the interaction with a 32-mer template.[74] The autocatalytic template replication of simpler organic polymeric compounds has been also achieved,[75] although the evolutionary connection of this alternative system and the biochemistry operating in current organisms cannot be postulated.

A further insight into the replication processes leading to LUCA comes from the top-down approach. The presence of DNA (carrying the genetic information) and proteins (performing metabolic functions) in all extant cells poses a catch-22 like paradox: DNA is required to produce proteins, while DNA replication cannot occur in the absence of proteins. At present, it is known that RNA can perform both functions: it still acts as genomic material in some viruses and all viroids, and, beginning in the early 1980s of the last century,[76] it has become clear that some RNA molecules (termed ribozymes) are able to perform catalytic functions in current organisms. Seven classes of natural ribozymes catalyze the cleavage or ligation of RNA: group I and group II autocatalytic introns, RNase P, hairpin, hammerhead, hepatitis delta virus (HDV) ribozyme and Varkud satellite ribozyme.[77]

The eighth class is the peptidyl transferase center of the ribosome, which catalyzes the formation of a peptide bond between two amino acids during translation.[78]

The evidence at hand led to the proposal that (as already suggested by Woese, Crick and Orgel in the 1960s) there might have been an RNA world prior to the establishment of the current DNA/RNA/protein world.[79] During the RNA world, genotype and phenotype (mainly represented by DNA and proteins, respectively, in extant organisms) should have been combined in a single type of macromolecule: RNA. The development of *in vitro* selection techniques in 1990 by means of the so-called SELEX method[80,81] has increased the repertoire of ribozymes, thus unveiling the functional plasticity of nucleic acids and supporting the plausibility of an RNA world. Nevertheless, it is currently impossible to postulate complex metabolisms based exclusively on ribozymes. Moreover, although template-dependent RNA polymerase ribozymes have been evolved *in vitro* with progressively better performance (and steadily decreasing the high mutation rate that affects replication in those systems[82,83]), we are still far from envisaging a ribozyme which is able to catalyze its own replication, an essential feature for the evolvability of RNA-based protocells.

Additional physico-chemical constraints (as the limited prebiotic abundance of ribonucleotides and the low stability of RNA in solution) might have hindered the *de novo* establishment of an RNA world. Therefore, different polymers analogous to nucleic acids have been postulated to have preceded RNA at the early stages of the evolution of genetic molecules, thus constituting putative "pre-RNA worlds".[84] These artificial analogues include glycerol-derived nucleic acid (GNA), threose nucleic acid (TNA), locked nucleic acid (LNA) and pyranosyl-RNA (pRNA), as well as a molecule with peptidomimetic backbone: peptide nucleic acid (PNA).

5. Theoretical Approaches to Replication and Metabolism

Solving the puzzle of the origins of life requires explaining the advent of template replication, metabolism and membrane-based compartmentalization. Replicating systems and metabolic networks might have appeared independent to each other. However, at a certain point they must have combined in the way to a protocell. Before the first protocell appeared as a unit of selection, it is almost certain that replicating molecules were loosely bound to their local system, such that the exchange of chemical information was frequent. LUCA was a late product, highly complex and evolved, of molecular evolution. At an intermediate stage, protocells likely originated in an age of promiscuous mix among genes, where the horizontal

exchange of hereditary information was the rule. Regarding metabolism, Darwin's seminal ideas, Oparin's and Haldane's hypotheses, and Miller–Urey's experiment favored a heterotrophic origin of life that has been further pursued by most theoretical models. Simple biomodules are just the starting point of this precellular world.

5.1. *Template replication and Darwinian evolution*

Experimental evidence at hand supports that short RNA (or RNA-like) polymers of random or quasi-random sequence could have been produced and accumulated abiotically. In these pools of polymers some biochemical functions (including the activity of simple ribozymes) could have been present. Therefore, basic RNA functionality could have preceded RNA-catalyzed template replication. At the base of this possibility lies the peculiar sequence-structure-function relationship in RNA. An important property of polymeric biomolecules is that they fold in tridimensional structures following thermodynamic rules. The native structure of biopolymers (in particular, RNA and proteins) is critical in the definition of their biochemical function. A key point often disregarded in models of evolution of replicators is the redundancy of the sequence-structure map. A clear example is provided by RNA. In short RNA oligomers (up to 40 nucleotides) of random sequence, the most abundant secondary structures are of the hairpin and stem-loop types.[17] Interestingly, it is known that certain RNA molecules with hairpin structure can display RNA ligase activity[85] and, thus, they could promote the concatenation of random polymers and the generation of modular, larger molecules even in the absence of template replication. When the huge number of different RNA sequences folding into the same structure is taken into account, the appearance of catalytic function is no longer a formidable obstacle.[18,19]

The existence of many different genotypes (RNA sequences) causing the same phenotype (their structure and, eventually, function) is supported by a series of more or less direct evidences. A number of experiments with functional RNA molecules have been devised to prove that neutral networks of phenotypes (consisting of all genotypes with the same phenotype) are not only ubiquitous, but in close contact in the space of genotypes. That is, almost any pair of different phenotypes can be retrieved by means of one or a few nucleotidic changes in an appropriate sequence. Schultes and Bartel[20] experimentally proved this fact by selecting two evolutionary unrelated ribozymes of almost equal length (89 nucleotides): one was a (synthetic) class III ligase ribozyme that catalyzes RNA ligation; the other was a (natural) HDV ribozyme that catalyzes cleavage and assists in the replication of the viral genomic RNA. It was necessary to change

about 40 nucleotides in each of the original sequences to generate an intersection sequence able to fold into each of the original structures and perform with notable competence the catalytic activity of both molecules under either fold. All intermediate steps between the original sequences and the intersection were catalytically active at levels mostly comparable to the initial ribozymes, and sometimes even better. This was a solid demonstration of the sequence-function redundancy and a clear evidence that dramatic alterations in structure and function are a few mutations away in sequence space.[21] Actually, a low fidelity of replication could have been advantageous to maintain variability (even within the same phenotype) and to enhance the adaptation of early molecular populations.[22] Thus, we must conclude that the naif relationship one sequence-one function is over-demanding and should not be indiscriminately applied, especially to short RNA sequences. In particular when modelling molecular quasispecies (as discussed below), attention should be focused on the phenotype, and models should systematically assume that a given function, in general, can be retrieved from a huge number of different sequences (genotypes) without *a priori* homology.

A second important aspect of molecular evolution before enzymatic template replication set in is the functional form of the growth rate of the abundance of chemical species (e.g. RNA oligomers). If the limiting step in the production of a certain molecular type is random polymerization, it will accumulate at most linearly in time. Experiments of non-enzymatic template replication yield faster growth rates, but they are still sub-exponential.[23] Any sub-exponential increase in the abundance of a population permits coexistence of species.[24] Differences in growth rates in linear or parabolic growth, for example, are not sufficient to outcompete other species present. In those plausible early scenarios, thus, survival of the fittest was not yet possible. Instead, different species could have accumulated in the environment and a high chemical diversity was maintained. That molecular heterogeneity was probably advantageous to foster the emergence of an eventually complex biochemistry, and the encapsulation of genetic molecules together with low molecular weight chemical species in reproducing vesicles could have aided in the selection process.[25]

Most models dealing with replicator dynamics assume that chemical species grow exponentially (although parabolic and hyperbolic growths may have been common in some scenarios[25]). A classical model is Eigen's quasispecies,[26] where the effect of frequent mutations was first analyzed. Manfred Eigen considered a number of chemical species (polymers), each characterized by a specific sequence and a particular replication rate, and affected by the same error rate of replication. One of the key results of the

model was that the concentration of the molecular species replicating at the fastest rate vanishes when the error rate of replication approaches a critical (finite) value which is of the order of the inverse of the polymer length. Beyond that threshold, genetic information cannot be maintained and the quasispecies enters into the so-called "error catastrophe". The limitations derived from Eigen's model are substantially alleviated when phenotypes, instead of genotypes, are the target of selection. Phenotypes can be maintained with remarkably higher error rates, as has been demonstrated with ribozymes.[27] Further, the precise value at which a given phenotype disappears depends on the characteristics of the phenotype itself,[28] particularly on the number of sequences representing it (i.e., on the size of its neutral network).[29]

At present, combined theoretical and empirical evidence suggest that a replication fidelity of 10^{-3} mutations per nucleotide and round of copy could be enough for RNA molecules of size 7000–8000 nucleotides to be maintained. This value is close to the typical genome length of most RNA viruses, and at the verge of the amount of genetic information hypothetically required to sustain a minimal cell.[25] Still, the obstacles found in the abiotic appearance of replicating molecules, including the problem of insufficient replication fidelity, has led to the proposal of alternative systems, among them hypercycles and compositional ensembles.

5.2. *Catalytic cycles*

A hypercycle is a population of molecules that interact by aiding in each other's replication.[30] For a hypercycle to be viable, each species has to aid replicating the following one, closing in circle. Competition between replicators is substituted in this scenario by cooperation between species, eventually maintaining a higher amount of genetic information distributed among shorter molecules. The hypercycle is a particular case of catalytic cycle where the growth of the species occurs at a hyperbolic rate. The self-organization of simple chemical species into catalytic cycles, however, also meets profound difficulties, and has been considered highly implausible.[31]

Autocatalytic networks of interacting proteins were proposed long ago.[32] Assuming a finite probability for a polypeptide to catalyze a chemical reaction involving other proteins in the ensemble, it was suggested that a self-sustained autocatalytic network would unavoidably emerge in a diverse enough population of proteins. An obstacle to the feasibility of such networks is however their evolvability.[35,37] On the experimental side, no large catalytic cycle has been produced so far, though small cycles have been chemically engineered. Current chemical reaction networks, embedded

in larger systems, cannot work properly without enzymes due to the high specificity of reactions and to energetic balance requirements.[33] A step-by-step construction of a large cycle, with selection of viable sub-networks at each stage draws a more plausible path than the spontaneous emergence of a catalytic cycle in full.[34] New species and reactions would be slowly incorporated, and the addition of small cofactors first, and efficient catalysts later, could perhaps prevent the proliferation of side reactions.

Theoretical advances have arrived by establishing the formal conditions that give plausibility to the appearance of a especially relevant type of catalytic cycles: reflexively autocatalytic and *F*-generated (RAF) sets. RAF sets capture the idea of catalytic closure, that is of a self-sustaining set supported by a steady supply of (simple) molecules from some reservoir.[36] It has been demonstrated that RAF cycles have indeed a high probability of appearance, even if the involved species display modest catalytic activity. RAF sets can be divided into a number of connected autocatalytic cores which can function as units of heritable adaptations in reaction networks.[37] This requires that more than one chemical reaction network be encapsulated into a compartment to allow competition and selection of networks.[38] The two elements together (plausibility of spontaneous appearance and limited evolvability) bring catalytic cycles back to the stage of important elements in the long way leading to the first protocell.

An often discussed problem to obtain stable catalytic cycles is the appearance and taking over of parasitic species. Several mechanisms have been proposed to counteract the deleterious effect of parasites, among them the evolution of catalytic cycles on a surface.[39] This spatial restriction could have helped as well in the selection of increased replication accuracy.[40] There is another possibility to limit the action of parasites that relies on group selection, as proposed in the stochastic corrector model.[16]

5.3. *Compositional ensembles: the lipid world*

The ease with which simple amphiphilic molecules and lipids of various kinds spontaneously self-assemble to form micelles and vesicles has led to the proposal of a "lipid world".[41] If some particular lipid composition could enhance the incorporation of further molecules, vesicles would grow autocatalytically and eventually divide due to simple physical forces, as discussed above. This would lead to the selection of autocatalytic, fast growing vesicles, in front of other possible compositional ensembles. A major criticism to the lipid world is that it lacks a sufficient capacity for evolvability. The simplicity of the underlying chemistry strongly limits the number of different phenotypes, understood as possible different combinations of lipid species, and in consequence hereditary variation is

severely limited.[42] In addition, it was shown that the replication of those ensembles is so inaccurate, that the "fittest" variant cannot be maintained in the population.[43]

6. Paths Towards Protocells

Most authors would agree that an evolutionarily relevant protocell should contain a metabolic subsystem, replicating molecules carrying the heritable information, and a semi-permeable boundary which is able to keep those components together.[44] Further, the protocell should be able to divide and distribute the chemical species between daughter protocells. As minimal living organisms, they have to behave as autopoietic systems.[45] In the sections above we have discussed different systems that consider only part of those essential elements. The conceptual integration of all of them is feasible,[86] and the experimental combination of every pair of elements leading to binary subsystems (template-boundary, metabolism-template and boundary-metabolism) has been tackled over the last decade, as a step forward towards the bottom-up construction of ternary, full-fledged biological systems. In particular, the combination of replicating nucleic acid polymers and reproducing compartments has been partially accomplished.[46] Among the various efforts aimed at combining metabolic networks and membrane compartments, a successful approach has allowed the encapsulation of the sugar synthesizing formose reaction into lipid vesicles.[87]

Current efforts towards the direct construction of ternary systems try to develop artificial cells formed by a lipidic membrane and an informational polymer able to replicate within.[47–49] Major issues are the exchange of nutrients and waste products with the extracellular environment as well as the growth, division and evolvability of such artificial cells,[50] before more ambitious goals can be attempted. Additionally, most of these approaches, though highly valuable from a biotechnological viewpoint, rely on the use of complex mixtures of compounds derived from extant organisms (typically, either *E. coli* extracts or large collections of gene products). Those semi-synthetic approaches cannot be strictly considered as relevant evolutionary paths towards protocells, and their usefulness in the field of the origins of life seems limited.[59] In the development of an artificial cell, the information derived from systems biology approaches could be also of relevance.[51]

Theoretical integrative scenarios, though by no means trivial, are ahead of experiments for obvious reasons. An early proposal of protocell is Ganti's Chemoton,[52] a closed system which grows as a consequence of its internal metabolism (represented by an autocatalytic network), has a bilayer

membrane formed by a molecule produced by the metabolism, and contains a replicating molecule carrying genetic information. Life is characterized by the fact that metabolism operates out of thermodynamic equilibrium, taking energy from its environment and using it to generate biomolecules. Thus, any realistic model of a protocell should take energetic balance into account.[53] Using knowledge on extant cells, additional important details have been added to the models, as a molecule that could act as energetic currency or particular membrane components able to generate chemiosmotic gradients.[54] However, the large formal complexity of models designed under this approach, which usually need a large number of kinetic equations for all chemicals involved, forbids to obtain general principles. The complex dynamics of the many coupled equations make those systems, in general, sensitive to endogenous and environmental fluctuations. Eventually, it is difficult to assess their robustness and evolvability. This nonetheless, these complex models contribute to the identification of the conditions that a viable protocell should verify, and to the understanding of the causal construction of extant-like cells from minimal metabolisms. Simpler formal models have highlighted specific limitations or requirements that protocells have to fulfill to be self-sustainable, for instance, how minority molecules limit the overall growth rate of protocells, requiring that the reproduction of the cell and the replication of the genetic molecule be synchronized.[55]

Research on protocells and minimal cells, provided that it takes into account evolutionary constraints, nowadays plays a pivotal role in the study of the origins of life. As Luisi and coworkers have stated: *"(...) there has been an abrupt rise of interest in the minimal cell. It appears that one additional reason for this rise of interest lies in a diffused sense of confidence that the minimal cell is indeed an experimentally accessible target."*[50] This is the feeling that all researchers on the origin of life probably share. The quest for the principles and mechanisms allowing the transition from inorganic to living matter would yield a scientific reward. Answering that question could deeply change our understanding of ourselves, our relationship with any other living being, and, foreseeably, our view of all the inanimate matter that conforms our universe: it may be just life-to-be.

Acknowledgments

The authors acknowledge the support of Spanish MICINN through projects EUI2008-00158, BIO2010-20696 and FIS2011-27569, as well as of Comunidad de Madrid through project MODELICO (S2009/ESP-1691).

References

1. M. Gargaud, Ph. Claeys, P. López-García, H. Martin, Th. Montmerle, R. Pascal and J. Reisse (eds)., *From Suns to Life* (Springer, Netherlands, 2006).
2. A. C. Forster and G. M. Church, *Mol. Syst. Biol.* **2**, 45 (2006).
3. S. L. Miller, *Science* **117**, 528 (1953).
4. A. P. Johnson, H. J. Cleaves, J. P. Dworkin, D. P. Glavin, A. Lazcano and J. L. Bada, *Science* **322**, 404 (2008).
5. D. Fox, *Sci. Am.* (28 March, 2007).
6. D. K. Kondepudi, R. J. Kaurman and N. Singh, *Science* **250**, 975 (1990).
7. F. C. Frank, *Biochim. Biophys. Acta* **11**, 459 (1953).
8. K. Soai, T. Shibata, H. Morioka and K. Choji, *Nature* **378**, 767 (1995).
9. Y. Saito and H. Hyuga, *J. Phys. Soc. Jpn.* **73**, 1685 (2004).
10. C. Viedma, *Phys. Rev. Lett.* **94**, 065504 (2005).
11. M. W. Powner, G. Gerland and J. D. Sutherland, *Nature* **459**, 239 (2009).
12. J. M. Edmond, K. L. von Damm, R. E. McDuff and I. Measures, *Nature* **297**, 187 (1982).
13. J. B. Corliss, J. Baross and S. E. Hoffman, *Oceanol. Acta* **4**, 59 (1981).
14. W. Martin, J. Baross, D. Kelley and M. J. Russell, *Nat. Rev. Microbiol.* **6**, 805 (2008).
15. D. Braun and A. Libchaber, *Phys. Biol.* **1**, 1 (2004).
16. J. M. Smith and E. Szathmáry, *The Major Transitions in Evolution* (Oxford University Press, New York, 1997).
17. M. Stich, C. Briones and S. C. Manrubia, *J. Theor. Biol.* **252**, 750 (2008).
18. S. C. Manrubia and C. Briones, *RNA* **13**, 97 (2007).
19. C. Briones, M. Stich and S. C. Manrubia, *RNA* **15**, 743 (2009).
20. E. A. Schultes and D. P. Bartel, *Science* **289**, 448 (2000).
21. G. F. Joyce, *Science* **289**, 401 (2000).
22. E. J. Hayden, E. Ferrada and A. Wagner, *Nature* **474**, 92 (2011).
23. G. von Kiedrowski, *Angew. Chem. Int. Ed.* **25**, 932 (1986).
24. E. Szathmáry and I. Gladkih, *J. Theor. Biol.* **138**, 55 (1989).
25. E. Szathmáry, *Philos. Trans. R. Soc. B* **361**, 1761 (2006).
26. M. Eigen, *Naturwissenschaften* **58**, 465 (1971).
27. Á. Kun, M. Santos and E. Szathmáry, *Nat. Genet.* **37**, 1008 (2005).
28. N. Takeuchi, P. H. Poorthuis and P. Hogeweg, *BMC Evol. Biol.* **5**, 9 (2005).
29. M. Stich and S. C. Manrubia, *J. Theor. Biol.* **280**, 117 (2011).
30. M. Eigen and P. Shuster, *Naturwisseschaften* **64**, 541 (1977).
31. L. Orgel, *Proc. Natl. Acad. Sci. USA* **97**, 12503 (2000).
32. S. Kauffman, *J. Theor. Biol.* **22**, 437 (1969).
33. P. Schuster, *Orig. Life Evol. Biosph.* **40**, 407 (2010).
34. S. Jain and S. Krishna, *Proc. Natl. Acad. Sci. USA* **98**, 543 (2001).
35. S. Lifson, *J. Mol. Evol.* **44**, 1 (1997).
36. W. Hordijk, J. Hein and M. Steel, *Entropy* **12**, 1733 (2010).
37. V. Vasas, Ch. Fernando, M. Santos, S. Kauffman and E. Szathmáry, *Biol. Direct* (2011).

38. G. Wachtershauser, *Microbiol. Revs.* **52**, 452 (1988).
39. S. C. Manrubia and J. F. Poyatos, *Europhys. Lett.* **64**, 557 (2003).
40. P. Szabo, I. Scheuring, T. Czaran and E. Szathmáry, *Nature* **420**, 340 (2002).
41. D. Segré, D. Ben-Eli, D. W. Deamer and D. Lancet, *Orig. Life Evol. Biosph.* **31**, 119–145 (2001).
42. E. Szathmáry, *Philos Trans. R. Soc. London B* **361**, 1669 (2000).
43. V. Vasas, E. Szathmáry and M. Santos, *Proc. Natl. Acad. Sci. USA* **107**, 1470 (2010).
44. T. Ganti, *The principles of life* (Oxford University Press, Oxford, 2003).
45. F. J. Varela, H. R. Maturana and R. Uribe, *Biosystems* **5**, 187 (1974).
46. J. P. Schrum, T. F. Zhu and J. W. Szostak, *Cold Spring Harb. Perspect. Biol.* **2**, a002212 (2010).
47. S. Rasmussen, L. Chen, D. Deamer, D. C. Krakauer, N. H. Packard, P. F. Stadler and M. A. Bedau, *Science* **303**, 963 (2004).
48. S. S. Mansy, J. P. Schrum, M. Krishnamurthy, S. Tobé, D. A. Treco and J. W. Szostak, *Nature* **454**, 122 (2008).
49. V. Noireaux, J. T. Maeda and A. Libchaber, *Proc. Natl. Acad. Sci. USA* **108**, 3473 (2011).
50. P. L. Luisi, F. Ferri and P. Stano, *Naturwissenschaften* **93**, 1 (2006).
51. V. De Lorenzo and A. Danchin, *EMBO Reps.* **9**, 822 (2008).
52. T. Ganti, *BioSystems* **7**, 15 (1975).
53. K. Ruiz-Mirazo and A. Moreno, *Artif. Life* **10**, 235 (2004).
54. F. Olasagasti, A. Moreno, J. Peretó and F. Morán, *Bull. Math. Biol.* **69**, 1423 (2007).
55. A. Kamimura and K. Kaneko, *Phys. Rev. Lett.* **105**, 268103 (2010).
56. W. Schopf and B.M. Packer, *Science* **237**, 70 (1987).
57. C. R. Woese, O. Kandler and M. L. Wheelis, *Proc. Natl. Acad. Sci. USA* **87**, 4576 (1990).
58. A. Moya, R. Gil, A. Latorre, J. Peretó, M. P. Garcillán-Barcia and F. de la Cruz, *FEMS Microbiol. Rev.* **33**, 225 (2009).
59. K. Ruiz-Mirazo, C. Briones and A. de la Escosura, *Chem. Rev.*, in press.
60. H. J. Kim, A. Ricardo, H. I. Illangkoon, M. J. Kim, M. A. Carrigan, F. Frye and S. A. Benner, *J. Am. Chem. Soc.* **133**, 9457 (2011).
61. A. Ricardo, M. A. Carrigan, A. N. Olcott and S. A. Benner, *Science* **303**, 196 (2004).
62. J. Oró, *Biochem. Biophys. Res. Commun.* **2**, 407 (1960).
63. R. M. Hazen and D. A. Sverjensky, *Cold Spring Harb. Perspect. Biol.* **2**, a002162 (2010).
64. D. G. Blackmond, *Phil. Trans. R. Soc. London B* **366**, 2878 (2011).
65. J. Bujdák and B. M. Rode, *Orig. Life Evol. Biosph.* **29**, 451 (1999).
66. R. Liu and L. E. Orgel, *Orig. Life Evol. Biosph.* **28**, 47–60 (1998).
67. B. M. Rode, D. Fitz and T. Jakschitz, *Chem. Biodivers.* **4**, 2674 (2007).
68. H. R. Kricheldorf, *Angew. Chem., Int. Ed.* **45**, 5752 (2006).
69. J. P. Ferris, *Phil. Trans. R. Soc. London B* **361**, 1777 (2006).
70. S. Rajamani, A. Vlassov, S. Benner, A. Coombs, F. Olasagasti and D. Deamer, *Orig. Life Evol. Biosph.* **38**, 57 (2008).

71. J. L. Bada and A. Lazcano, *Science* **296**, 1982 (2002).
72. S. E. Maurer, D. W. Deamer, J. M. Boncella and P. A. Monnard, *Astrobiol.* **9**, 979 (2009).
73. P. Stano and P. L. Luisi, *Chem. Commun. (Cambridge)* **46**, 3639 (2010).
74. D. H. Lee, J. R. Granja, J. A. Martinez, K. Severin and M. R. Ghadiri, *Nature* **382**, 525 (1996).
75. T. Tjivikua, P. Ballester and J. Rebek, Jr., *J. Am. Chem. Soc.* **112**, 1249 (1990).
76. T. R. Cech, A. J. Zaug and P. J. Grabowsky, *Cell* **27**, 487 (1981).
77. E. Puerta-Fernández, C. Romero-López, A. Barroso-delJesus and A. Berzal-Herranz, *FEMS Microbiol. Rev.* **27**, 75 (2003).
78. P. Nissen, J. Hansen, N. Ban, P. B. Moore and T. A. Steitz, *Science* **289**, 920 (2000).
79. W. Gilbert, *Nature* **319**, 618 (1986).
80. C. Tuerk and L. Gold, *Science* **249**, 505 (1990).
81. A. D. Ellington and J. W. Szostak, *Nature* **346**, 818 (1990).
82. W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner and D. P. Bartel, *Science* **292**, 1319 (2001).
83. A. Wochner, J. Attwater, A. Coulson and P. Holliger, *Science* **332**, 209 (2011).
84. G. F. Joyce, *Nature* **418**, 214 (2002).
85. J. M. Buzayan, W. L. Gerlach and G. Bruening, *Nature* **323**, 349 (1986).
86. E. Szathmáry, M. Santos and C. Fernando, *Top. Curr. Chem.* **259**, 167 (2005).
87. P. M. Gardner, K. Winzer and B. G. Davis, *Nat. Chem.* **1**, 377 (2009).
88. F. J. Dyson, *Origins of Life* (Cambridge University Press, Cambridge, 1985).