

## 5. Prebiotic Chemistry – Biochemistry – Emergence of Life (4.4–2 Ga)

ROBERT PASCAL and LAURENT BOITEAU

*Département de Chimie, Université Montpellier II, Montpellier, France*  
(E-mails: [rpascal@univ-montp2.fr](mailto:rpascal@univ-montp2.fr); [laurent.boiteau@univ-montp2.fr](mailto:laurent.boiteau@univ-montp2.fr))

PATRICK FORTERRE

*Institut de Génétique et Microbiologie, Université Paris-Sud, Orsay, France*  
(E-mail: [fortere@igmors.u-psud.fr](mailto:fortere@igmors.u-psud.fr))

MURIEL GARGAUD

*Observatoire Aquitain des Sciences de l'Univers, Université Bordeaux1, Bordeaux, France*  
(E-mail: [gargaud@obs.u-bordeaux1.fr](mailto:gargaud@obs.u-bordeaux1.fr))

ANTONIO LAZCANO

*Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico*  
(E-mail: [alar@correo.unam.mx](mailto:alar@correo.unam.mx))

PURIFICACIÓN LÓPEZ-GARCÍA and DAVID MOREIRA

*Unité d'Ecologie, Systématique et Evolution, Université Paris-Sud, Orsay, France*  
(E-mails: [puri.lopez@ese.u-psud.fr](mailto:puri.lopez@ese.u-psud.fr); [david.moreira@ese.u-psud.fr](mailto:david.moreira@ese.u-psud.fr))

MARIE-CHRISTINE MAUREL

*Institut Jacques Monod, Université Paris 6, Paris, France*  
(E-mail: [marie-chistine.maurel@ijm.jussieu.fr](mailto:marie-chistine.maurel@ijm.jussieu.fr))

JULI PERETÓ

*Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, València, Spain*  
(E-mail: [Juli.Pereto@uv.es](mailto:Juli.Pereto@uv.es))

DANIEL PRIEUR

*Laboratoire de Microbiologie des Environnements Extrêmes, Université de Bretagne  
Occidentale, Brest, France*  
(E-mail: [Daniel.Prieur@univ-brest.fr](mailto:Daniel.Prieur@univ-brest.fr))

JACQUES REISSE

*Faculté des Sciences Appliquées (CP 165/64), Université Libre de Bruxelles, Brussels, Belgium*  
(E-mail: [jreisse@mach.ulb.ac.be](mailto:jreisse@mach.ulb.ac.be))

(Received 1 February 2006; Accepted 4 April 2006)

**Abstract.** This chapter is devoted to a discussion about the difficulties and even the impossibility to date the events that occurred during the transition from non-living matter to the first living cells. Nevertheless, the attempts to devise plausible scenarios accounting for the emergence of the main molecular devices and processes found in biology are presented including the role of nucleotides at early stages (RNA world). On the other hand, hypotheses on the development of early metabolisms, com-

partments and genetic encoding are also discussed in relation with their role in extant living organisms. The nature of the Last Common Ancestor is also presented as well as hypotheses on the evolution of viruses. The following sections constitute a collection of independent articles providing a general overview of these aspects.

**Keywords:** Origin of life, chemical evolution, origin of genetic information, metabolism, membrane

### 5.1. A Word of Caution about Chronology

JACQUES REISSE, LAURENT BOITEAU, PATRICK FORTERRE, MURIEL GARGAUD, ANTONIO LAZCANO, PURIFICACIÓN LÓPEZ-GARCÍA, MARIE-CHRISTINE MAUREL, DAVID MOREIRA, ROBERT PASCAL, JULI PERETÓ, DANIEL PRIEUR

Compared to astronomers and geologists, chemists and biochemists find themselves in a very difficult situation when asked to participate to a collective work on the dating of significant events in astrobiology. Little information is available that can allow them to date in detail the events that took place when the protosolar nebula started to collapse and eventually the young Earth was formed and life first appeared in our planet. Dating the origin of the constituents of living matter is in itself a huge problem. Some of these molecules were probably already present in the interstellar cloud long before it splitted into various nebulae. One of these nebulae was the protosolar nebula, which therefore must have contained a vast ensemble of organic molecules. It is generally accepted that during the accretion of the solar system, these organic molecules were probably destroyed in the inner part of the system, but some of them may have remained intact or with little modifications in small volatile-rich bodies like comets or, perhaps, even in the parent bodies of carbonaceous chondrites. During post-accretional processes these preformed organic compounds, which may have included amino acids or nucleic bases, were delivered to the young Earth together with water molecules and other simple volatiles. Therefore, it could be argued that some of the components of living systems were probably formed before the solar system itself. Obviously, it can also be argued that the hydrogen atoms found in past and extant life forms were formed very soon after the Big Bang, and that the life story started many billions year before the origin of the solar system itself! The choice of an origin is *always* arbitrary and requires a careful definition of what is the starting point.

Although there are some dissenting views, it has been generally assumed that life appeared on Earth once the physico-chemical conditions of the primitive environment were compatible with the presence of organic polymers like nucleic acids and polypeptides. Most researchers agree that one of these conditions was the presence of liquid water, and that therefore the prebiotic stage may have started around 4.4–4.2 billions year ago. On the other hand,

although the identification of the oldest traces of life remains a contentious issue, it is generally agreed that a microbial biosphere had already developed on Earth 3.5 billions years ago. In between, chemical evolution took place. Any attempt to put exact dates on particular steps may be futile at the time being. Of course, and as we will see in the following pages, chemists and biochemists are able to make some reasonable assumptions about the sequence of some steps leading from the synthesis and accumulation of biochemical monomers to the first cells. Let us consider two examples: given the chemical lability of RNA, it is possible that oligopeptides may have existed before polyribonucleotides began to accumulate, and that RNA in turn evolved prior to DNA. Nevertheless, no one is able to state if the accumulation of a necessary (whatever “necessary” may have been) amount of polypeptides required 1 year or 1 million years. Similarly, nobody is ready to claim that the synthesis of a “sufficient” amount of polynucleotides took a “short time” or a “very long time”, considering that even “short” and “very long” cannot be defined!

Individually, chemical reactions can be fast or slow depending on the rate constants values (which themselves are strongly depending on the temperature, pH, ionic strength, and so on) but also on the reactant concentrations. Given our poor understanding of primitive Earth conditions, it could be argued that it is impossible to estimate, even roughly, the time necessary, in an unknown place on the young Earth to go through the various steps required for the emergence of a living cell. However, simulation experiments can provide important insights on the rate of chemical syntheses and/or degradations under various conditions.

In order to avoid the problem related to the degradation of organic molecules and of the supramolecular systems into which they may have evolved, it is tempting to suggest that the multi-steps syntheses which led to the formation of the first living cell were fast processes. It is difficult to estimate the rate of self-organization of the precursors of life into replicating systems, because the chemical steps are unknown. Whatever the time scale required for the appearance of an informational polymer, once formed it must have persisted at least long enough to allow its replication. If polymers formed by a slow addition of monomers, this process must have been rapid compared to rates of hydrolysis, especially if a considerable amount of genetic information was contained in the polymer. Self-replicating systems capable of undergoing Darwinian evolution must have emerged in a period shorter than the destruction rates of their components; even if the backbone of primitive genetic polymers was highly stable, the nitrogen bases themselves would decompose over long periods of time. In fact, it can be argued that the accumulation of all components of the primitive soup will be limited by destructive processes, including the pyrolysis of the organic compounds in the submarine vents (large amounts of the entire Earth's oceans circulate through the ridge crests every 10 million years facing temperatures of 350 °C or more).

However, nobody knows the number of aborted “attempts” before the final result was reached and life appeared and persisted. What is sure is the fact that presently, prebiotic chemistry is unable to put dates on steps between a system that is definitively non-living and a system which could be recognized as living. It is doubtful that the situation will change in the near future: as we said previously, our poor knowledge on the actual conditions of the primitive environment, the absence of molecular relics from the prebiotic organic world, the dependence of organic reactions rates on external conditions and, last but not least, the fast degradation of organic molecules and supramolecular organic assemblies will perhaps preclude for ever to date prebiotic events. The only thing that can be done (with great care!) is to suggest a plausible sequence of events. This is what the authors of the following chapter have tried to do.

Following this approach, chemical evolution and the first stages of biochemical evolution may be thought as a succession of stages corresponding to chemical or biochemical structures of increasing complexity. However, this view is probably biased by our current description of the basic properties of extant life. Then early chemical and biochemical stages may have followed one another gradually and probably involved the coexistence of a large number of pathways, the majority of which disappeared and might be considered as dead branches of the evolution tree. It is also possible that what we call the “RNA world” (or the additional hypotheses suggested by those that follow other alternatives) simply corresponded to a subset of molecules that we consider as *qualitatively* significant among a large number of systems that came to a dead end, so that these stages may have occurred simultaneously. For these reasons, the chapter on chemical and early biological evolution has been divided into several parts corresponding to structures of increasing complexity, which does not mean that evolution proceeded following this sequence: for instance, it will never be possible to know if the confinement within vesicles preceded or not the appearance of primitive replicating genetic polymers, or if they coexisted from the very beginning.

Evolutionary biologists interested in very early evolution are largely in front of similar problems to those faced by chemists and biochemists. They could be helped in principle by some imposed constraints coming from geology, micropaleontology, and geochemistry, including bona-fide microfossils, isotopic data and molecular fossils. However, sometimes these signatures are ambiguous and controversial (see chapter 7.1), and in most cases do not allow any inference about the lifestyle, the metabolism or the phylogenetic affiliation of the corresponding living species. Biologists are obliged to use indirect arguments based on what they know about modern microorganisms, comparing genes and proteins involved in metabolic pathways to suggest a plausible evolutionary route that explains their observed contemporary patterns of distribution in organisms and, eventually, a series of

possible sequential events towards the past that leads to speculate about the nature of the earliest metabolic strategies. However, in some cases this depends strongly on the model of evolutionary reconstruction that is chosen. One or a few parsimonious scenarios can be identified but, ultimately, evolution does not necessarily follow the most parsimonious way among all those that are possible, so that absolute certainty about the succession of early evolutionary steps is unattainable. Nobody knows for sure how the last unicellular common ancestor looked like or even when it became the dominant form of life on the young Earth. So, for different reasons, chemists interested in prebiotic chemistry and biologists interested in the first living organisms must accept that they have few things to say about chronology but they have to explain why it is so. The works that give the impression that the “how” and “when” questions concerning the origin of life are solved, except for few details, do not contribute to the development of astrobiology. It is much better to list the problems for which plausible explanations exist but also those which remain without solution. In these last cases, it is sometimes necessary to question the question and to try to find another way to formulate the questioning. In science, the impossibility to find a solution to a problem can be the proof that the problem is not well formulated.

## **5.2. A Scenario Starting from the First Chemical Building Blocks**

LAURENT BOITEAU, ROBERT PASCAL

Although chronology is impossible in prebiotic chemistry, building plausible scenarios, linking the possibilities of simple abiotic processes to what the most ancient biochemical pathways are supposed to be, is the central goal of this field of science. In other words, there is no definitive answer to the question: how long did it take for life to be present? But there may be an (or several) answer(s) to: what is the sequence of stages that were covered for living beings to emerge in the abiotic environment of the early Earth, where simple organic compounds were synthesized in the early atmosphere, in other places (hydrothermal vents, volcanic plumes) or delivered from the outer space by meteoritic bombardment? The object of prebiotic chemistry is then to solve a problem that depends on the definition of what was the environment of the early Earth (unsteady over the 4.4–2 Ga b.p. era assigned to the origin of life) and on the definition of what could have been the biochemistry of early living organisms inferred from the sciences of evolution (see part 5.7). It is then necessary to analyze what occurred in the black box separating the two stages to get a continuous sequence of incremental steps as usually observed in the evolution of living organisms, the state of the system at the entrance and at the exit being not clearly known.

### 5.2.1. AVAILABILITY OF ORGANIC MATTER AND ENERGY

The origin of organic matter and energy on the early Earth is still debated mainly because of the lack of precise indications establishing the composition of the atmosphere (and especially the relative content in  $H_2$  and  $CO_2$ , see chapter 6.2.1) and the surface and ocean temperatures (that may have been comprised within a range of less than  $0\text{ }^\circ\text{C}$  to above  $80\text{ }^\circ\text{C}$ , see chapters 4.4 and 6.2). Moreover, as pointed out by Miller (1998), the conditions under which organic molecules can be formed in an equilibrium state or in a stationary state (thanks to a flux of energy or of reactive species with a dehydration or oxidation state away from equilibrium) are usually those in which these molecules become instable. Then, any synthetic process in a given area (or under defined conditions) must also involve a quenching step to protect the activated mixture from further degradation by transferring it into a milder environment. This is true for any kind of energy sources including lightning, ultraviolet irradiation of the atmosphere, hydrothermal processes of synthesis and even for the delivery of exogenous organic matter by bombardment. Therefore, the claim of a hypothetical synthetic or deleterious character of a given energy source should only be considered after taking into account the efficiency of the quenching process. Anyway, it is unlikely that a single energy source may have been responsible for the formation of the organic matter needed by the origin of life process and the early stages of life evolution.

Two different kinds of processes have contributed to feed the early Earth with biogenic molecules, biomolecules and the building blocks needed for polymer synthesis: the exogenous delivery and the endogenous formation as a result of energy input in the atmosphere or in the oceans. The relative importance of the two processes was principally dependant on the degree of oxidation of the primitive atmosphere and thus on its hydrogen content (Chyba and Sagan, 1992; Chyba, 2005). The initial Miller experiment (Miller, 1953) was carried out in a mixture with a high content in  $H_2$ ,  $CH_4$ , and  $NH_3$  that is favourable for the synthesis of organic molecules. Then, it has been considered that the escape of hydrogen from the atmosphere to the outer space had been so fast that the Earth atmosphere rapidly reached a composition based on  $CO_2$  and  $N_2$  with a low content in  $H_2$  (Kasting, 1993), much less favourable for synthesis. As a result, the amount of biogenic molecules synthesized on the early Earth may have been much lower and life depended on extraterrestrial delivery. Indications have been reported supporting the formation of organic molecules in the solar nebula and then their delivery to the early Earth since a high organic content can be found in meteorites and comets (Mullie and Risse, 1987; Cronin and Chang, 1993). Then, it is obvious that the delivery of biogenic compounds from the outer space played a role at that stage but its relative contribution mainly depended on the local production, which can only be deduced from hypothetical

models of the early atmosphere. A recent model of the evolution of atmosphere supports a hydrogen escape that could have been much slower than previously believed (Tian et al., 2005), so that the amount of organic matter produced on Earth could have been sufficient for the emergence of life. It has also been proposed that hydrothermal synthesis, by which organic molecules could have been synthesized by heating taking advantage of the presence of minerals (see for example Holm and Andersson, 1998; Russell and Martin, 2004), may have contributed to endogenous production. The catalogue of molecules produced includes high-energy biogenic basic species such as HCN, cyanate, formaldehyde and other aldehydes (Miller, 1998). These simple molecules can undergo different processes under conditions simulating the primitive Earth capable of yielding the building blocks of biochemistry: amino acids (Strecker reaction of aldehydes), nucleic bases, and sugars (formose reaction from formaldehyde). But it is more difficult to define a scenario of chemical evolution by which the system became more complex allowing the formation of macromolecular and supramolecular components of life and their combination into metabolic processes.

### 5.2.2. FAVOURABLE AREAS FOR PREBIOTIC CHEMICAL PROCESSES

Since the knowledge of the Earth formation suggests that an habitable ocean may have been present as early as 4.4 Ga (see chapter 4.2.2), it is possible to speculate that life emerged within the first few hundred million years of the history of the planet and may have survived the Late Heavy Bombardment (see chapters 4.4 and 6.2.1). But we have no indication about the occurrence of this event. Hypothetical places that would have been favourable can be inferred from the processes leading to the accumulation of organic matter in a specific environment. For instance, as soon as emerged lands were present, it can be devised that ocean tides (induced by the vicinity of the Moon) may have allowed the formation of pools concentrating monomers and may have triggered wetting/drying cycles capable of inducing their polymerization possibly with the assistance of minerals (Rohlfing, 1976, Lahav et al., 1978, Rode et al. 1999). They may also have induced solid-gas reactions that have been shown to induce peptide bond formation (Commeyras et al., 2002). Volcanic areas, rich in reduced and sulphur-derived compounds, or hydrothermal vents in the ocean may be considered as favourable locations for these processes as well.

### 5.2.3. CHEMICAL EVOLUTION THROUGH A STEPWISE PROCESS

Prebiotic chemistry is customarily divided into different stages corresponding to an increasing degree of complexity of the entities involved:



- The synthesis of building blocks (amino acids, nucleic bases, nucleosides, nucleotides...).
- The formation of polymers (nucleic acids, peptides).
- The emergence of supramolecular architectures including the formation of the membrane and hence of individual cells.

This partition corresponds to one of the first hypothesis on the origin of life suggesting that life began in a “primordial soup” containing all the chemical components needed for feeding the first living organism. Since Miller’s experiment was performed, prebiotic chemistry has demonstrated the capacity of making a wide range of building blocks available (amino acids, carbohydrates, nucleic bases) under favourable abiotic environments (Miller, 1998). The next degree of organization, the formation of peptides and nucleic acids, may have been the result of interactions (involving possibly activating agents) in a pool of inactivated building blocks. Then, the origin of life is considered in this hypothesis as a sequence of events resulting from more or less improbable encounters of building blocks leading finally to a system capable of self-replication. As it does not involve any driving force, this hypothesis suggests that the origin of life was highly improbable. As a result, it would have been less unlikely starting from high concentrations of building blocks maintained during a long period of time in a stable environment required for monomers to accumulate and polymers to have a sufficient lifetime.

#### 5.2.4. CHEMICAL EVOLUTION THROUGH A DYNAMIC PROCESS

The partition of the origin-of-life problem into several stages corresponds to an approach that could rather be considered as based on the present day biochemistry way of thinking and teaching. This partition may however be misleading since the capacity to evolve and to promote the emergence of new properties must have been the most important feature of the chemical system from which life arose. This behaviour is observed for chemical systems maintained in states far away from equilibrium by a constant or erratic flux of matter and energy (Eigen, 1971; Nicolis and Prigogine, 1977). The origin of life may therefore be considered as the emergence of individual structures with new properties (and incorporating an information content capable of being reproduced) from a prebiotic network of chemical reactions linking high-energy species to inactivated products. Then the supply of energy and activated reactants may be considered as the driving force for chemical evolution. Additionally, the process is likely to be dependant on physico-chemical constraints governing the reactivity of activated biogenic compounds. Increasing the strength of constraints and of the driving force would have made the emergence of life less improbable so that an unstable environment may have been less deleterious than in a process purely governed by chance. In other words, these constraints may have influenced the rate of



evolution of the system and, consequently, the chronology of events even though the process remained non-deterministic and historically unique (as far as we know). The synthesis of building blocks, polymers and supramolecular structures may then have been associated in a single process so that no activation of monomers was needed, the energy needed for polymerization being carried by precursors. From high-energy biogenic compounds, any early chemometabolic pathways leading to more complex species required a sequence of kinetically and thermodynamically spontaneous reactions linking substrates to products (Weber, 2002). Catalytic abilities to overcome activation barriers were probably limited (Weber, 2002). Of course, thermodynamics would never by itself identify the pathway leading to the first replicating molecule or supramolecular edifice (Kuhn and Kuhn, 2003). But, as any other chemical reactions, chemometabolic pathways are governed by thermodynamic constraints, the quantitative analysis of which can be used to rule out or to validate these processes.

#### 5.2.5. CATALYTIC ACTIVITY AND INFORMATION STORAGE

A minimal form of life would have needed the association of a carrier of information content and the chemical activity needed for its replication. The form of life that we are presently familiar with on Earth mainly developed around two classes of biopolymers: protein and peptides carrying catalytic activity and nucleic acids carrying information. But noteworthy exceptions (ribozymes, nucleotide-like enzyme cofactors) have been considered as strong indications that the situation may have been quite different at early stages of evolution (see part 5.5). More generally, we have no indication that the different classes of biomolecules played the same role at early stages as in modern biochemistry. Three main hypotheses can be devised concerning the development of this process. Two of them correspond to a stepwise process in which one class of polymers could have acquired a replicating ability and then, in a later stage, the translation process was discovered leading to the modern protein-nucleic acid world. The advantage of this process from a chronological point of view is that it can be described by a sequence of stages involving an increasing complexity, but the driving force leading to the emergence of new properties is not obvious. On the contrary, the last process, corresponding to coevolution, may have developed from a probably quite complicated network of organic reactions, maintained far away from equilibrium through the feeding with activated biogenic compounds.

##### 5.2.5.1. *Emergence of life in a peptide world:*

Since there is no accepted abiotic pathways leading to nucleosides and nucleotides, and since amino acids are the more easily synthesized building blocks in prebiotic experiments and have been detected in extraterrestrial organic matter

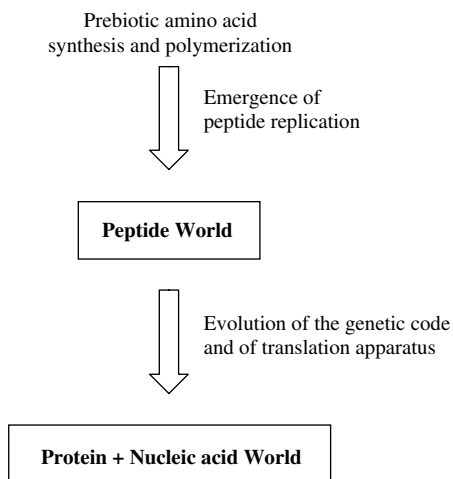


Figure 5.1. The sequence of stages corresponding to the origin of life from self-replicating peptides.

(meteorites), it is tempting to consider that life developed from a peptide only world. Peptide bonds can be formed thermodynamically from free amino acids by heating or under dehydrating conditions (generally in the presence of catalysts) or under the effect of activating agents. Additionally, these peptides may have been subject to hydrolysis so that elongation, by addition of activated monomers at the N-terminus, was taking place at the same time as peptide bonds were cleaved. These simultaneous processes may have led to some kind of selection through peptide protometabolisms (Commeyras et al., 2002; Huber et al., 2003, Plankensteiner et al., 2005). Peptides with sequences capable of catalytic activities may have been formed (Barbier and Brack, 1992) and others may have become prone to self-replication through a selection process in a population of continuously growing and disappearing random sequences. However, peptide self-replication is a highly improbable process since there are for instance  $10^{13}$  ( $20^{10}$ ) different decamers starting with the modern set of twenty amino acids. This hypothesis is unlikely because efficient catalysis usually requires peptides having several secondary structure domains ( $\alpha$ -helices or  $\beta$ -sheets) associated to each other to ensure a properly defined fold (Corey and Corey, 1996). This requirement is achievable only for peptides having a sufficient length<sup>1</sup> (ca. 50 residues) that need an encoding system for the sequence to be

<sup>1</sup> In this chapter, (usually short) random poly-amino acid sequences, which generally do not fold into definite structures, are called (poly) *peptides*. This is a major difference with *proteins*, the catalytic or recognition abilities of which are the result of stable three-dimensional structures determined by their genetically encoded sequences in living organisms (Fersht, 1999).

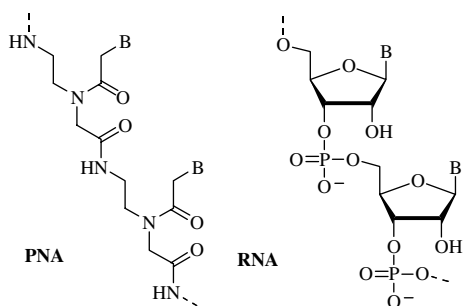


Figure 5.2. Peptide Nucleic Acid (PNA) as hypothetical early information carriers, their structure compared to that of RNA.

reproduced accurately from monomers. The next difficulty with this hypothesis is the need of a subsequent process that would have converted the amino acid sequences into genetic information. An alternative would be the existence of a first set of residues capable of carrying information such as Peptide Nucleic Acids involving a peptide-like backbone and nucleic bases on the side chains (Bohler et al., 1995; Nielsen, 1999; Nelson et al., 2000). In this way, peptides could have played both the role of information carriers and of catalysts in a pre-RNA world (Figure 5.2).

A replacement of this early genetic information system by the modern nucleic acid-based system is then needed. But the rationale for a radical change like this is not clear and it would require that no remnant of this former information storage system have been preserved by evolution.

#### 5.2.5.2. Emergence of life in an RNA world:

According to the RNA world hypothesis (Gesteland et al., 1999) nucleic acid played both the role of information storage and the role of catalysts at an initial or intermediate stage of evolution (see part 5.5). Although there is no consensus on the absence of peptides in an RNA world, their presence is usually considered as non-essential at that stage (Figure 5.3). However, the direct emergence of a self-replicating RNA sequence is usually considered as unlikely because there is presently no prebiotically plausible pathway for the synthesis of mononucleotides so that the RNA world may have been preceded by another system of replicating molecules called the pre-RNA world (Orgel, 2004). Moreover, there is no obvious driving force that would have led to the selection of the translation apparatus in an RNA world, though it has been suggested that RNA folding could have been improved by short coded peptides (Noller, 2004).

#### 5.2.5.3. Emergence of life from an RNA-peptide coevolution process:

The former hypotheses suppose inherently that most biochemical evidences of the initial stage have been lost since a switch of information support or of

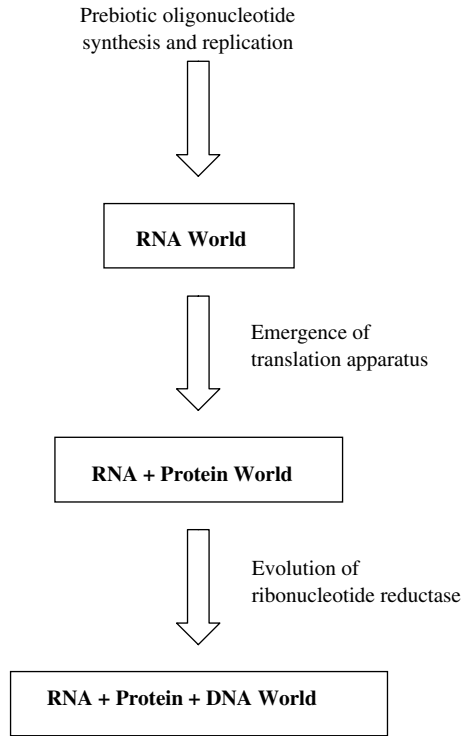


Figure 5.3. The sequence of stages corresponding to the direct emergence of an RNA world.

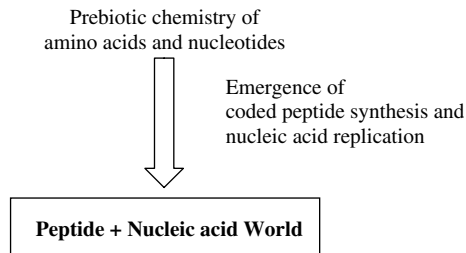


Figure 5.4. Coevolution process for the emergence of life leading directly to an RNA protein world.

catalytic molecules has erased records of the initial process. There is an alternative possibility that life evolved using directly both systems, which seems more complex on a first sight. But if life developed from an RNA–peptide coevolution process, the translation machinery could be considered as a metabolic remnant of the initial stage (Figure 5.4). Indeed, covalent bonds between amino acids and AMP or tRNAs are formed in the bio-

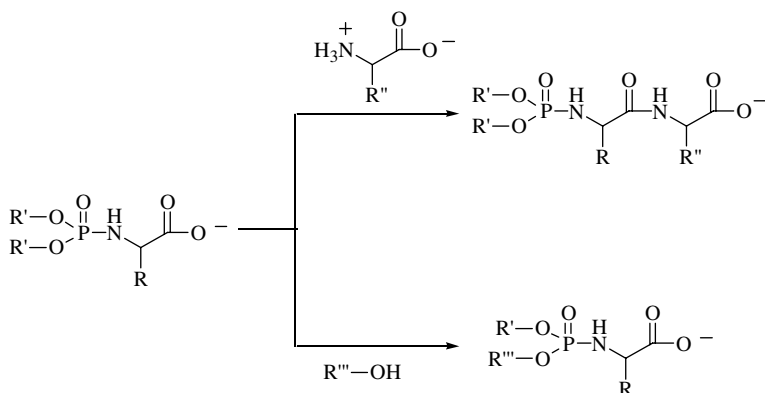


Figure 5.5. Ester exchange and peptide formation from *N*-(dialkylphosphoryl)amino acids.

chemical activation of amino acids and are essential in the reliability of the translation process. It is also supported by the catalytic activity of the ribosome (Nissen et al., 2000) on peptide bond formation.<sup>2</sup> Several chemical processes compatible with a coevolution scenario have been reported. For instance, it could have been involved as early as in the prebiotic synthesis of essential building blocks, which is supported by the discovery of a catalytic activity of amino acids (Weber, 2001) on the formose reaction, a likely process for the formation of sugars from formaldehyde. Actually, there is increasing evidence that amino acids and peptides are capable of very interesting stereoselective catalytic activities in the formation of carbon-carbon bonds through aldol-forming reaction, which may have been of importance for the emergence of homochirality (Pizzarello and Weber, 2004; Cordóva et al., 2005).

*N*-(dialkylphosphoryl)amino acids display interesting capabilities for the prebiotic syntheses of peptides and polynucleotides (Cheng et al., 2004). An intramolecular phosphoric-carboxylic mixed anhydride has been proposed to explain the specific behavior of *N*-(dialkylphosphoryl)amino acids that spontaneously give oligopeptides upon standing in various solvents (Figure 5.5). Peptide formation was accompanied by diester exchange on the phosphoryl group via a reaction that may have been useful for nucleotide ligation (the key step of RNA synthesis). Because of their abilities in both peptide and nucleic acid oligomerization, *N*-phosphoryl amino acids could have played an important role in prebiotic chemistry on condition that a plausible pathway of synthesis of these compounds has been made available.

<sup>2</sup> However, it has been recently determined that the rate increase brought about by the ribosome is consistent with a role of entropy trap so that it could be the result of binding both reacting tRNAs at the convenient position for reaction, without need for additional catalysis by the ribosomal RNA (Sievers et al., 2004).

The idea that life began in a system linking nucleic acid replication and genetically coded peptide synthesis has also been presented (Sutherland and Blackburn, 1997; Borsenberger et al., 2004), so that the two pathways may not be viewed as two separate processes. The aminoacylation of RNA may have arisen from an other purpose and was then subverted by protein synthesis; this may explain why the translation process developed whereas any advantage from translation requires a reasonably full set of specifically aminoacylated tRNAs (Sutherland and Blackburn, 1997). Experimental support in favour of a linkage of amino acid and nucleotide chemistries at early stages is the reaction of amino acid *N*-carboxyanhydrides with inorganic phosphate (Biron and Pascal, 2004) and with nucleotides (Biron et al., 2005), leading to mixed anhydrides (Figure 5.6). This is the first demonstrated abiotic pathway that may have led to the most activated forms of amino acids found in biology, aminoacyl phosphates and aminoacyl adenylates, which are involved in ribosomal (Arnez and Moras, 2003) and non-ribosomal peptide syntheses (Marahiel et al., 1997; Healy et al., 2000). Since there is increasing evidence that NCAs can be considered as unexpectedly common prebiotic molecules (Taillades et al., 1999; Maurel and Orgel, 2000; Leman et al., 2004), it is possible that, at early stages, amino acid activation was not dependent on the energy provided by phosphoanhydrides. Since aminoacyl adenylates and amino acid anhydrides with other nucleotides are mixed anhydrides, they could be considered as activated nucleotides as well as activated amino acids; i.e. the activation of nucleotides at that stage may have been dependent on amino acid chemistry (Pascal et al., 2005).

A coevolution process is also supported by the behavior of 3'-phosphonucleosides (Biron et al., 2005) that undergo two different intramolecular reactions with NCAs through the mixed anhydride leading either to a cyclic phosphodiester or to an amino acid ester that is reminiscent of aminoacylated tRNA (Figure 5.7).

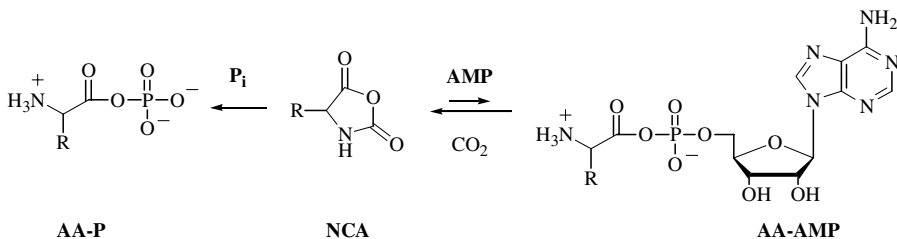


Figure 5.6. The formation of aminoacyl phosphates (AA-P) and aminoacyl adenylates (AA-AMP) from amino acid *N*-carboxyanhydrides and inorganic phosphate ( $P_i$ ) or adenosine-5'-monophosphate (AMP), respectively.

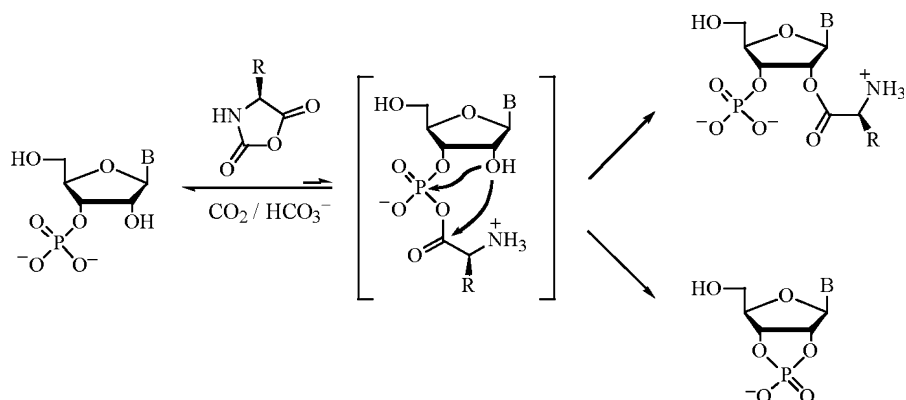


Figure 5.7. Intramolecular reactions of mixed anhydride derived from 3'-phosphorylated nucleosides.

### 5.3. Hypothesis about Early Metabolisms

DANIEL PRIEUR

All organisms living on Earth to day (with the exception of viruses, see part 5.8) are organized on a cellular basis. The cell is the fundamental unit of living matter, and is an entity isolated from its environment by a membrane (mostly made of lipids). Within this membrane are gathered molecules and sub-cellular structures required for cell life, and particularly macromolecules such as proteins, lipids, polysaccharides and nucleic acids. A cell may be single or associated to others, forming tissues, organs and finally complex organisms. But in any case, a living cell possesses five major functions: metabolism, growth (reproduction), differentiation, chemical communication and evolution. Consequently, the universal common ancestor of all living organisms (LUCA, see part 5.7), was almost probably also organized on a cellular basis.

The first cell function listed above is metabolism, which is the sum of all biochemical reactions occurring within the cell. These reactions aim to synthesize macromolecules (anabolism) or to obtain the energy required for all cellular functions (catabolism). A quick look at the biochemical pathways described for the metabolisms used by contemporary cells, and particularly prokaryotic cells, shows that these pathways (for catabolism and anabolism) are very complicated in the sense they require a variety of transporters, electron carriers, enzymes, co-enzymes, and a complex genome encoding all the required proteins. The first entities certainly used a rather simple process to gain their energy from their environment, and consequently to synthesize their components. No traces, no signatures, no dates are known for these early metabolisms, and the biologist can only imagine something simple, and



compatible with the conditions existing for the period preceding the first records of microbial fossils (see chapter 7.1).

### 5.3.1. PRELIMINARY DEFINITIONS

Altogether, contemporary living organisms use a variety of metabolic pathways to gain the energy required by their living functions. Although several energy sources such as magnetism and thermal gradients have been theoretically considered (Schulze-Makuch and Irwin, 2002), all present living organisms depend on chemical or photochemical reactions for their energy (Madigan et al., 2003).

Living organisms can be classified in several groups, depending on their energetic metabolism. Those using light as energy source are called *phototrophs*. Those gaining their energy from chemical reactions are named *chemotrophs*. In this case, those using organic molecules as energy sources are *chemo-organotrophs*, while those using inorganic molecules as energy sources are called *chemo-lithotrophs*.

In order to build their components, cells uptake in their environments small amounts of micro-nutrients (Cr, Co, Cu, Mn, Mo, Ni, etc) but larger amounts of macro-nutrients (C, N, H, O, P, S, K, Mg, Ca, Na, Fe). Among them, carbon is particularly important since it is present in all macromolecules. Cells using organic carbon molecules for biosynthesis are named *heterotrophs*, while those using inorganic carbon (carbon dioxide) are called *autotrophs*.

### 5.3.2. HOW THE FIRST ENTITIES PRESUMABLY GAINED THEIR ENERGY AND CARBON?

Despite the report of microfossils showing prokaryotic morphotypes in very old rocks (Westall et al., 2001), the metabolism displayed by these organisms cannot be deduced from these observations. If a photosynthetic (most probably anoxygenic) metabolism (based on morphological similarities with extant stromatolites) has been suggested for putative microorganisms occurring in stromatolite fossils aged of 3.5 Ga or younger (Schopf, 1993), this is not finally proved (Brasier et al., 2002), and the question of metabolism is still open for other old microfossils reported from Barberton rocks in Australia (see chapter 7.1).

The scenario given by Madigan et al. (2003) is rather convincing and explained below.

Whatever the exact dating for the first cellular-like entities with an energetic metabolism, it is obvious that this event occurred under anoxic conditions. From our knowledge of energetic metabolisms used by contemporary prokaryotes, several anaerobic metabolisms can be hypothesized: anoxic photosynthesis, fermentation and anaerobic respiration. All require a variety

of enzymes, electron carriers, and for phototrophs, photosynthetic pigments, all involved in rather complex pathways that one cannot easily imagine for early energy generating systems. What kind of simple mechanism could be considered? If anoxic photosynthesis is excluded because of its complex pathways and because photosensitive pigments are already evolved molecules, a simple chemolithotrophy must be considered seriously. On the primitive Earth, likewise nowadays, a variety of reduced inorganic molecules (putative electron donors) did exist, and among them, molecular hydrogen represents an excellent candidate. Molecular hydrogen is a common energy source for prokaryotes living in geothermal (terrestrial and marine) areas (Prieur, 2005), and it has been demonstrated that this compound could drive hyperthermophilic ecosystems in Yellowstone National Park (Nealson, 2005; Spear et al., 2005). Molecular hydrogen may be a product of interactions between hydrogen sulphide and ferrous iron, or between protons and ferrous iron in the presence of UV radiation as an energy source (Spear et al., 2005). Whatever its origin, molecular hydrogen belongs with protons to a redox couple whose reduction potential ( $E_0' = -0.42$  V) is very favourable for electron donation. With such an electron donor, there is a wide choice of putative electron acceptors in the absence of molecular oxygen. Among those (still inorganic) used by extant prokaryotes, elemental sulphur ( $S^0$ ) represents a good candidate. Elemental sulphur and hydrogen sulphide form a redox couple whose reduction potential ( $E_0' = -0.28$ ) is favourable for free energy generation, but does not require a long series of electron carriers. As shown on Figure 5.8, a primitive hydrogenase would have been the single enzyme

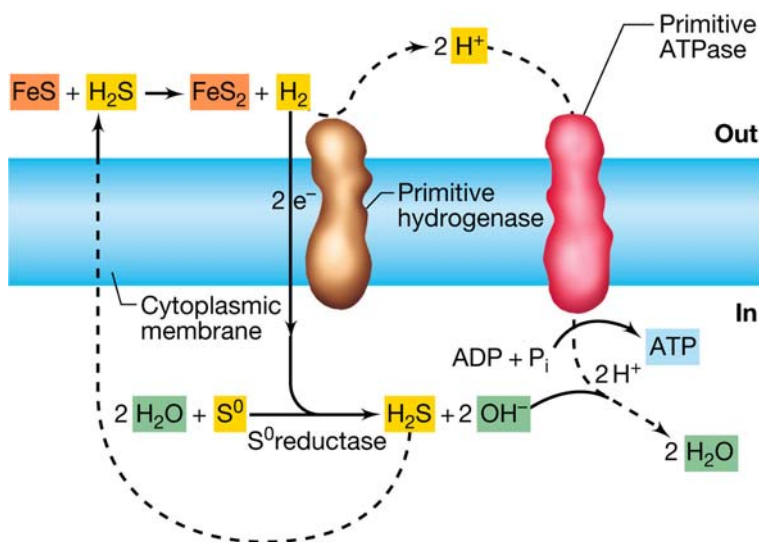


Figure 5.8. A hypothetical energy-generating metabolism for primitive cells (Madigan et al., 2003).

required for uptaking molecular hydrogen, producing protons and electrons, and consequently forming a proton gradient and a proton motive force. A primitive ATPase (this enzyme is present for all living organisms) would have used this proton motive force for ATP generation. This hypothesis is rather convincing, taking into account that the mechanism hypothesized is rather simple and requires only two primitive enzymes, and that the inorganic compounds involved were most probably abundant on the primitive Earth.

If such an energy generating mechanism is suggested, one must now consider the question of the carbon source (s).

An organism which uses molecular hydrogen as electron donor and elemental sulphur as an electron acceptor is called a chemolithotroph carrying out anaerobic respiration (see definitions above). Such organisms are, for most of those living nowadays, also autotrophs and use carbon dioxide as a single carbon source. Autotrophy is a property of various organisms that do not use organic carbon for their energy generation: anoxygenic and oxygenic photosynthesizers and chemolithotrophs. To transform carbon dioxide into organic carbon, they use rather complex pathways such as the Calvin cycle, but also the reverse citric acid or the hydroxypropionate cycles. These pathways require a variety of enzymes, which are again difficult to imagine working all together in primitive organisms. Since organic compounds (whatever their origins) existed on the primitive Earth, it is more probable that the primitive chemolithotrophic cells utilized already formed organic carbon as carbon sources, in the same way as extant mixotrophic organisms (such as certain sulphur-oxidizers) are doing today. If that was the case, although it is more difficult to consider it as a common and unique energy generating mechanism, a kind of simple fermentation<sup>3</sup> could also be considered as a possible hypothesis. Actually, various fermentations are nowadays supporting the growth of many different organisms. Although they generate rather low amounts of energy compared with respirations using electron donors and acceptors with a great difference of reduction potentials, they cannot be eliminated as a possible early metabolism. But it is difficult to imagine which particular fermentation among many possibilities, would have been the first. Consequently, molecular hydrogen might remain the most probable energy generating compound for early living organisms, as it is still today used by a variety of mesophilic and thermophilic Bacteria and Archea (Prieur, 2005).

<sup>3</sup> In fermentation an organic compounds serves as electron donor and carbon source, an electron acceptor is temporarily generated from an intermediate compounds resulting from the degradation of the initial carbon and an energy source.

## 5.4. Origin and Evolution of Compartments

PURIFICACIÓN LÓPEZ-GARCÍA, DAVID MOREIRA AND JULI PERETÓ

Although impossible to date, and hard to place in a relative succession of events, compartments must have appeared very early during the emergence of life to enclose metabolism (self-maintenance) and information storage in entities that reproduced and could undergo natural selection. For many authors, the earliest life forms must have had boundaries: life only appeared when the state of self-reproducing compartments was reached (Varela et al., 1974; Morowitz et al., 1988; Deamer, 1997; Luisi, 1998; Peretó, 2005). Contemporary cells are surrounded by membranes that assure their integrity, facilitate the necessary exchange with the external environment (diffusion of gases, active transport of ions and metabolites) and harbour energy-transducing systems that take advantage of the ion gradients maintained across the membrane using primary energy sources (i.e. visible light or chemical reactions). The exploitation of electrochemical gradients across membranes to supply energy (the chemiosmotic theory) is a universal property of living cells. Any hypothesis on the origin of life must explicitly state the way to convert energy into complex structures and organization, something that is only achieved in terrestrial life through compartments (Harold, 2001). Cell membranes are essentially made out of phospholipids, amphiphilic molecules composed of a hydrophilic glycerol-phosphate head bound to long hydrophobic fatty acid or isoprenoid tails. They organise in bilayers and host different proteins involved in transport and energy-transducing processes (Figure 5.9). Of course, at early stages, compartments must have been defined by much simpler barriers and less functions, possibly only two: definition of a 'self' boundary and of a not-quite-impermeable one, i.e. one allowing the import of ions and metabolites but retaining most of the internally produced material. A remarkable *ab initio* problem would be the osmotic crisis generated by the enclosure of polymers. This situation could be mitigated by the coevolution of membranes and primitive ion exchange systems that necessarily could not be sophisticated protein structures but simplest molecular devices (e.g. the *Escherichia coli* non-proteinaceous calcium channel made out of polyphosphate and polyhydroxybutyrate; Reusch, 2000).

### 5.4.1. AMPHIPHILIC VERSUS NON-AMPHIPHILIC COMPARTMENTS

What was the nature of those early compartments? The fact that amphiphilic lipid bilayers are universal today suggests that protocellular compartments may have been founded on simpler molecular grounds with similar properties such as long single-chain monocarboxylic acids, alcohols or monoglycerides,

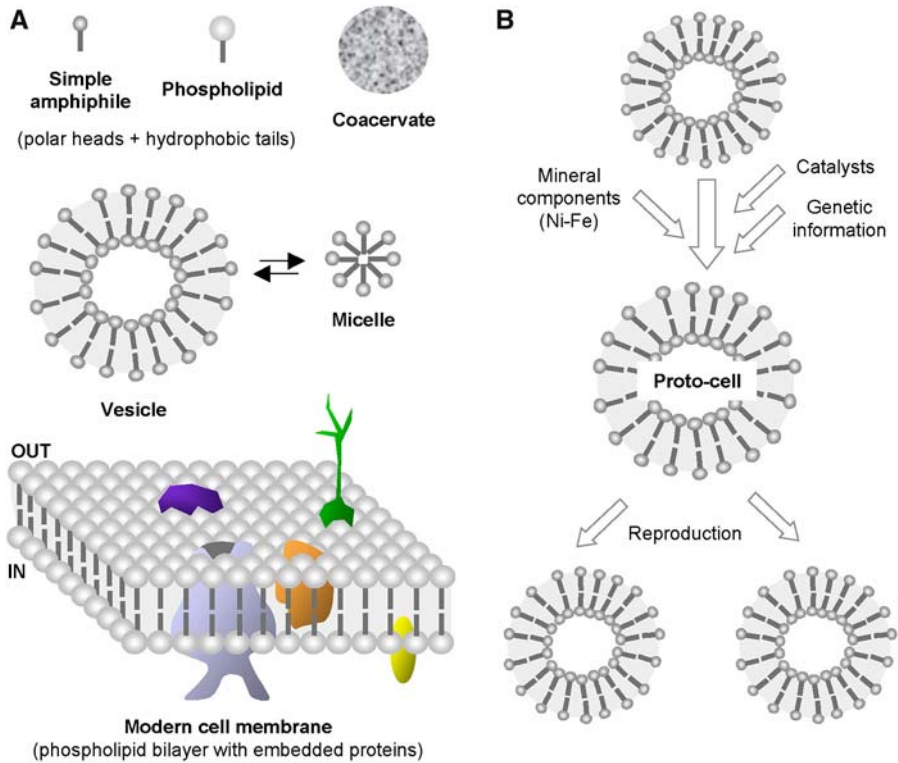


Figure 5.9. A. Various types of compartments and their components. B. A possible model of membrane evolution.

capable of self-assembly to form vesicles and available in prebiotic times (Deamer, 1986; Deamer et al., 2002; Monnard and Deamer, 2002). Oparin's coacervates (Oparin et al., 1976), spherical aggregates of macromolecular components, and micelles (Figure 5.9) are unlikely to have played a role in the origin of protocellular compartments. Amphiphilic vesicles are compatible with "prebiotic-soup" models, but they are also compatible with models of surface-metabolism (Wächtershäuser, 1988b) operating in early times. Surface metabolism likely played a role in the synthesis and accumulation of complex organic molecules (Lazcano, 2001; Monnard and Deamer, 2002) and perhaps in vesicle replication (Hanczyc and Szostak, 2004). Nevertheless, surface-based hypotheses are 'acellular' by definition. This inconvenience has been overcome by a proposal suggesting that the first compartments were tridimensional iron monosulphide bubbles that grew in hydrothermal environments, and that these mineral membranes persisted for a long time in actual biological evolution (Russell and Hall, 1997; Martin and Russell, 2003). Although such mineral compartments may have played also a role as initial chemical reactors, their persistence in relatively modern cells is highly

improbable (see part 5.7). Additional evidence comes from contemporary membranes, which do not form *de novo* but grow and divide from pre-existing membranes. This membrane heredity view (Cavalier-Smith, 2001) together with the heuristic principle of continuity (Morowitz et al., 1991) would conform to the idea that some type of amphiphilic bilayer existed since early times and that a kind of continuum allowed its evolution to date. Within this framework, amphiphiles and vesicles would be key chemical intermediates during life emergence (Ourisson and Nakatani, 1994; Luisi et al., 1999; Segré et al., 2001).

#### 5.4.2. SELF-ASSEMBLY AND EARLY EVOLUTION OF COMPARTMENTS

How and when did early compartments evolve? A possibility is that the earliest protocellular compartments resulted from the self-assembly of organic mixtures that were available on the planet. Long ( $C_{16-18}$ ) fatty acids and alcohols assemble spontaneously, forming vesicles above a concentration threshold. Shorter fatty acids ( $C_{14}$ ) would be even better candidates because they were easily synthesized and the bilayers formed are much more permeable, an advantage in times when membrane transport proteins had not yet evolved (Deamer, 1986; Monnard and Deamer, 2002). Interestingly enough, a selective passive incorporation of ribose into fatty acid vesicles has been shown (Sacerdote and Szostak, 2005), suggesting that preferential ribose uptake by primitive cells would play a role in the rise of a hypothetical RNA world. Long monocarboxylic acids were present in prebiotic conditions both by exogenous delivery and endogenous synthesis. The exogenous delivery is attested by the presence of mixed aliphatic and aromatic compounds in meteorites such as Murchison that, indeed, are able to form vesicles spontaneously (Monnard and Deamer, 2002). This kind of compounds can also be synthesized by Fischer-Tropsch-type reactions (Deamer et al., 2002). Deamer (1997) suggested that prebiotic conditions favoured saturated over unsaturated acids and considered that isoprenoid-type molecules were difficult to make abiotically. Interestingly, the presence of “impurities”, for instance pyrrolic compounds (the first pigments) increases vesicle permeability (Deamer, 1997). Furthermore, it has been shown experimentally that the interaction and incorporation of minerals (montmorillonite) to fatty acids catalyze vesicle formation (Hanczyc et al., 2003). Such vesicles can grow by incorporating surrounding fatty acids and divide (Hanczyc et al., 2003). Moreover, the growth of vesicles can generate incipient ion gradients (Chen and Szostak, 2004). RNA and other macromolecular species can be encapsulated in vesicles leading to an experimental approach to the simplest life forms (Luisi et al., 1994; Chen et al., 2004). Therefore, it is plausible that prebiotic mixed short-chain fatty acids formed vesicles that

encapsulated, sequentially, catalytic species and genetic systems to become self-reproducing compartments. Nevertheless, whether this occurred and how it occurred remains a matter of speculation. Attempts to re-create life *in vitro* by encapsulating macromolecules such as ribozymes (see part 5.5) and other catalytic species in replicative vesicles will certainly yield interesting results (Szostak et al., 2001). Hopefully, these experiments will at least contribute to test the feasibility of those hypotheses.

### 5.5. The Hypothesis of an RNA World

MARIE-CHRISTINE MAUREL

The time necessary to go from an habitable Earth to a protocell-like pro-caryote can be divided in three periods. During the first period molecular organics sources, from which building blocks of life could have appeared, accumulated on the early Earth. The next period led to macromolecular synthesis from small monomers and to the first metabolic steps including the formation of the first replicating polymers. Subsequently a scenario can be described for the development from random polymers to a replicative system, capable of evolving by mutation and natural selection. This last period, called the *RNA world period*, would have opened the door to evolutionary biology as we know it today, leading to organized and complex systems.

The question of how long did it take to go from prebiotic building blocks to the first living cell must be addressed according to the inherent constraints imposed by primeval conditions. Any extrapolations from results obtained in the laboratory to what may have occurred 4 billion years ago are tenuous. As a result we have to study the stability of all components in extreme conditions, that is the behaviour of monomers and macromolecules of life at high and cold temperatures (Schwartzman and Lineweaver, 2004; Vergne et al., 2006), with and without salt (Tehei et al., 2002), at low and high pH (Kühne and Joyce, 2003), at low and high pressure (Tobé et al., 2005; Di Giulio, 2005), in different redox conditions, radio-ionizing and cosmic conditions, solvent conditions etc., as well as in conditions simulating an ocean-boiling asteroid impact...

Diverse molecular ecosystems could potentially have arisen in these physico-chemical specifications and we have to take them into account especially if the hereditary criteria are retained as mandatory in designing life. Thus the half-lives for the decomposition of the components of life (amino acids, peptides, sugars, lipids) and of the first genetic materials, that are nucleobases must be measured and considered on the geological time scale. Also the balance between synthesis and degradation must lead to consistent concentrations.



Minerals and mineral surfaces, salt and crystals may help to stabilize macromolecules and monomers (Tehei et al., 2002; Cornée et al., 2004; Ricardo et al., 2004). Purines and pyrimidines have been found in sediment cores from both ocean and lake basins, some dating back as far as  $25 \times 10^6$  year, but they may be the result of contamination or decomposition under anhydrous conditions. On the other hand, adenine, uracil, guanine, xanthine, hypoxanthine nitrogenous bases and several organics have been found in the Murchison meteorite (Stoks and Schwartz, 1982) and it is now possible to detect subpicomoles of purine bases trapped in a mineral or a colloidal supports (El Amri et al., 2003, 2004, 2005).

Finally, these materials, if applicable to any origin of life theory founded on Darwinian evolution, may have resisted to several extinctions where the survival of a single organism (in a micro-environment) would be sufficient to reestablish an entire ecosystem.

Speculations from results obtained in the laboratory, specially unconstrained sequences obtained by *in vitro* selection, to what may have occurred 4 billion years ago, are weak. Again, a valuable approach lies in the examination and the experimental test of the resistance of RNA molecules under the inherent constraints imposed by prebiotic geochemical and geophysical conditions.

Lastly as it seems likely that the RNA World may not have been the pristine nucleic acid-dominated ecosystem but simply a transient go-between during the evolution to the contemporary DNA–protein world, considerations above would apply to any alternative pre-RNA backbone before the emergence of standard ribonucleic acids.

#### 5.5.1. THE RNA WORLD SCRIPT

A scenario of evolution postulates that an ancestral molecular world, *the RNA World*, existed originally before the contemporary DNA–RNA–Protein world meaning that the functional properties of nucleic acids and proteins as we see them today would have been produced by molecules of ribonucleic acids (Gilbert, 1986; Benner et al., 1989, 1993; Joyce, 1989; Orgel, 1989; Bartel and Unrau, 1999; Gesteland et al., 1999; Joyce and Orgel, 1999; McGinness et al., 2002; Joyce, 2002). RNAs occupy a pivotal role in the cell metabolism of all living organisms and several biochemical observations resulting from the study of contemporary metabolism should be stressed. For instance, throughout its life cycle, the cell produces the deoxyribonucleotides required for the synthesis of DNA from ribonucleotides, the monomers of RNA. Thymine, a base specific of DNA, is obtained by transformation (methylation) of uracil a base specific of RNA, and RNAs serve as obligatory primers during DNA synthesis. Finally, the demonstration that RNAs act as catalysts is an additional argument in favour of the presence during evolution

of RNAs before DNA. Therefore it seems highly likely that RNA arose before DNA during biochemical evolution, and for this reason DNA is sometimes considered as modified RNA better suited for the conservation of genetic information. This genetic privilege would constitute a logical step in an evolutionary process during which other molecules could have preceded RNA and transmitted genetic information. The idea of an *RNA world* rests primarily on three fundamental hypotheses, developed by Joyce and Orgel (1999):

- during a certain period in evolution, genetic continuity was assured by RNA replication;
- replication was based on Watson–Crick type base pairing;
- early catalysis was performed by small non genetically coded peptides and by ribozymes.

Orgel and his coworkers showed that starting from activated monomers, it is possible in certain conditions to copy a large number of oligonucleotide sequences containing one or two different nucleotides in the absence of enzyme (Inoue and Orgel, 1983; Joyce and Orgel, 1986; Orgel, 1992; Hill et al., 1993). On the other hand, Ferris and his coworkers studied the assembly of RNA oligomers on the surface of montmorillonite (Ferris, 1987; Ferris and Ertem, 1992). Thus, experimental results demonstrated that minerals which serve as adsorbing surfaces and as catalysts (Paecht-Horowitz et al., 1970; Ferris et al., 1996), can lead to accumulation of long oligonucleotides, given that activated monomers are available. One can thus envisage that activated mononucleotides assembled into oligomers on the montmorillonite surface or on an equivalent mineral surface. The longest strands, serving as templates, direct the synthesis of complementary strands starting from monomers or short oligomers, leading double-stranded RNA molecules to accumulate. Finally, a double RNA helix – of which one strand is endowed with RNA polymerase activity –, would dissociate to copy the complementary strand and to produce a second polymerase that would copy the first to produce a second complementary strand, and so forth. The RNA world would thus have emerged from a mixture of activated nucleotides. However, a mixture of activated nucleotides would need to have been available! Finally, when either the first replicative molecule, the template or one of its elements (nucleotides) is to be synthesized from the original building blocks, in particular the sugars that are constituents of nucleotides, a certain number of difficulties are encountered (Sutherland and Whitfield, 1997). First, synthesis of sugars from formaldehyde produces a complex mixture in which ribose is in low amounts. Second, production of a nucleoside from a base and a sugar leads to numerous isomers, and no synthesis of pyrimidine nucleosides has so far been achieved in prebiotic conditions. Finally, phosphorylation of nucleosides also tends to produce complex mixtures (Ferris, 1987). Consequently, onset of nucleic acid replication is

nearly inconceivable if one does not envisage a simpler mechanism for the prebiotic synthesis of nucleotides. Eschenmoser succeeded in producing 2,4-diphosphate ribose during a potentially prebiotic reaction between glycol aldehyde monophosphate and formaldehyde (Eschenmoser, 1999). It is thus possible that direct prebiotic nucleotide synthesis occurred by an alternative chemical pathway. Nevertheless, it is more likely that a certain organized form of chemistry preceded the RNA world, hence the notion of “genetic take-over”. Since the ribose-phosphate skeleton is theoretically not indispensable for the transfer of genetic information, it is logical to propose that a simpler replication system would have appeared before the RNA molecule. During the evolutionary process, a first genetic material, mineral in nature would have been replaced by another totally distinct material of organic nature. The hypothesis of a precursor of nucleic acid (Cairns-Smith, 1966, 1982; Joyce et al., 1987) is a relatively ancient idea, but it is only within the last few years that research has been oriented towards the study of simpler molecules than present day RNAs, yet capable of auto-replication. In the *Peptide Nucleic Acids* (PNA) of Nielsen and coworkers, the ribofuranose-phosphate skeleton is replaced by a polyamidic skeleton on which purine and pyrimidine bases are grafted (see Figure 5.2). Indeed, PNAs form very stable double helices with an RNA or a complementary DNA (Egholm et al., 1993) and can serve as template for the synthesis of RNA, or vice versa (Schmidt et al., 1997). Moreover, PNA–DNA chimeras containing two types of monomers have been produced on DNA or PNA templates (Koppitz et al., 1998). Eventually, the information can be transferred from PNAs (achiral monomers) to RNA during directed synthesis; the double helical molecule with a single complementary RNA strand is stable. Transition from a “PNA world” to an “RNA world” is hence possible. The group of Eschenmoser recently replaced the ribose moiety of RNA by a four-carbon sugar, threose, whose prebiotic synthesis seems easier. The resulting oligonucleotides designated TNAs, (3' → 2')- $\alpha$ -L-threose nucleic acid, can form a double helix with RNA (Schöning et al., 2000). TNA is capable of antiparallel Watson–Crick pairing with complementary DNA, RNA and TNA oligonucleotides. Finally, this leads us to a major conclusion, namely that a transition may have occurred between two different systems without loss of information.

From the point of view of evolution, the studies described previously demonstrate that other molecules capable of transmitting hereditary information may have preceded our present day nucleic acids. This is what Cairns-Smith coined the “take-over” (Cairns-Smith, 1982), the evolutionary encroachment or genetic take-over, or to some extent what François Jacob (1970) calls genetic tinkering, in other words, making new material from the old.

The role of cofactors at all steps of the metabolism and their distribution within contemporary groups of organisms suggests that a great variety of nucleotides was present in the ancestor common to all forms of life and before. Several authors have underscored the possible presence of coenzymes before the appearance of the translation machinery (White, 1976). Present-day coenzymes, indispensable cofactors for many proteins, would be living fossils of catalysts of primitive metabolism. Most coenzymes are nucleotides (NAD, NADP, FAD, coenzyme A, ATP...) or contain heterocyclic nitrogen bases and it is even possible to consider that catalytic groups that were part of nucleic enzymes were incorporated in specific amino acids rather than being "retained" as coenzymes. This could be the case of imidazole, the functional group of histidine, whose present synthesis in the cell is triggered by a nucleotide.

Work has been carried out based on the demonstration of esterase activity in a nucleoside analogue N<sup>6</sup>-ribosyladenine (Fuller et al., 1972; Maurel and Ninio, 1987; Maurel, 1992). This activity which is due to the presence of an imidazolyl group that is free and available for catalysis, is comparable to that of histidine placed in the same conditions. We have studied the kinetic behaviour of this type of catalyst (Ricard et al., 1996) and have shown that the catalytic effect increases greatly when the catalytic element, pseudohistidine, is placed in a favourable environment within a macromolecule (Décout et al., 1995). Moreover, primitive nucleotides were not necessarily restricted to the standard nucleotides encountered today, and because of their replicative and catalytic properties, the N<sup>6</sup> and N<sup>3</sup> substituted derivatives of purines could have constituted essential links between the nucleic acid world and the protein world.

### 5.5.2. THE CASE OF ADENINE

Purine nucleotides, and in particular those containing adenine, participate in a large variety of cellular biochemical processes (Neunlist et al., 1987; Maurel and Décout, 1999; Nissen et al 2000). Also, the ease with which purine bases are formed in prebiotic conditions (Oró, 1960) suggests that these bases were probably essential components of an early genetic system. Furthermore, purines have also been found in the Murchison meteorite showing the range of resistance of this molecule. The first genetic system was probably capable of forming base pairs of the Watson–Crick type, Hoogsteen and other atypical associations, by hydrogen bonds as they still appear today in RNA. It probably contained a different skeleton from that of RNA, and no doubt it also modified bases, thereby adding chemical functions, but also hydrophobic groups, and functions such as amine, thiol, imidazole, etc. Wächtershäuser (1988a) also suggested novel pairings of the purine–purine type.

Today, from a vast combination of nucleic acids, one can isolate aptamers that possess catalytic properties (RNA ligation, cleavage or synthesis of a peptide bond, transfer of an aminoacyl group, etc.). The first nucleic acids could possess independent domains, separated by flexible segments, creating reversible conformational motifs, dependent on ions and bound ligands. Thus, a 10 amino acid-long peptide can recognize fine structural differences within a micro RNA helix (discrimination can be made between two closely-placed microhelices). Just as protein and antibodies, RNA molecules can present hollows, cavities, or slits that make these specific molecular recognitions possible. RNAs must “behave as proteins”. Whatever the chronology and the order of appearance of the various classes of molecules, the importance lies in the shape, the scaffolding and the architecture that have allowed functional associations.

Starting from a heterogenous population of RNAs with  $10^{15}$  variants (a population of  $10^{15}$  different molecules), five populations of RNAs capable of specifically recognizing adenine after about ten generations have been selected (Meli et al., 2002). When cloned, sequenced and modelled, the best one among the individuals of these populations, has a shape reminiscent of a claw capable of grasping adenine. Is it the exact copy of a primitive ribo-organism that feeds on prebiotic adenine in prebiotic conditions? Functional and structural studies presently under way will highlight other activities, other conformations...

Following this line of investigation two adenine-dependent ribozymes capable of triggering reversible cleavage reactions have been selected. One of them is also active with imidazole alone. This result leads to very important perspectives (Meli et al., 2003).

A considerable amount of research has been focused on the selection of ribozymes *in vitro*. Recently, it was demonstrated that a ribozyme is capable of continuous evolution, adding successively up to 3 nucleotides to the initial molecule (McGinness, 2002). It is also possible to construct a ribozyme with only two different nucleotides, 2,6-diaminopurine and uracil (Reader and Joyce, 2002).

### 5.5.3. PROVISIONAL CONCLUSIONS

Very little is known to date about the behavior of macromolecules in “extreme” environments. How do structures behave? What are the major modifications observed? What are the conditions of structural and functional stability? How are the dynamics of the macromolecules and their interactions affected? What are the possibilities of conserving biological macromolecules in very ancient soils or in meteorites? Can we find traces of these macromolecules as molecular biosignatures, and if so in what form (Maurel and Zaccai, 2001; Tehei et al., 2002)?

The selection of thermohalophilic aptamers, RNAs resistant to high temperatures (80 °C) in the presence of salt (halites 30 million years old) (Vergne et al., 2002, 2006), will maybe allow us to answer some of these questions, that are fundamental for the search and the date of past traces of life, and of life on other planets...

## 5.6. The RNA/DNA Transition and the Origin of the Genetic Code

### 5.6.1. THE ORIGIN OF THE GENETIC CODE

JULI PERETÓ

The hypothesis of an RNA world is widely accepted (see part 5.2 and part 5.5). Nevertheless, the different evolutionary paths emerging from such a scenario have attracted much less attention. Did the protein synthesis and the genetic code emerge in an RNA-peptide world and, afterwards, DNA was invented in an RNA/protein world? Or did DNA precede proteins? All these and other alternative possibilities have been critically discussed by Dworkin et al. (2003). Albeit a precise chronology of the different transitions is impossible, we can try to establish the most parsimonious order of events based in our chemical and biochemical knowledge. Therefore, and although we are conscious that alternative solutions can not be totally ruled out, the current consensus around a growing experimental evidence favours an emergence of the machinery for biosynthesis of coded peptides in a world of ribozymes helped by amino acids and short peptides as cofactors (i.e. an RNA-peptide coevolution process, see part 5.2; Szathmáry, 1999; Noller, 2004). The view of an early emergence – in any case, before the universal ancestor, see part 5.7 – of the machinery for both transcription and translation is consistent with genomic and structural studies of the major macromolecular components of RNA polymerases and ribosomes (Fox and Naik, 2004).

#### 5.6.1.1. *Origin of protein synthesis in an RNA world*

The biosynthesis of coded proteins in modern cells is performed by the ribosome, an extraordinarily complex assembly of proteins and RNA whose high-resolution structure has been recently elucidated – the large subunit described by Ban et al. (2000), the small subunit by Wimberly et al. (2000), and the complete ribosome by Yusupov et al. (2001). There are good structural (Nissen et al., 2000; Hoang and Noller, 2004) and chemical (Zhang and Cech, 1997) reasons to propose that peptidyltransferase activity resides in the 23S rRNA of the large ribosomal subunit, although this view is not fully consistent with some biochemical and genetic observations (see Fox and

Naik, 2004, and references therein). It is widely assumed that the ribozymic nature of the ribosomal peptidyltransferase is a molecular fossil from the RNA world.

Protein biosynthesis is one of the more complex, energy dependent, metabolic processes. To perform its function, the ribosome is assisted by many different molecular components: tRNAs – which both activate amino acids through an ester bond and carry the anticodon triplet complementary of a codon triplet in the mRNA –, aminoacyl-tRNA synthases – aaRS, which catalyze the synthesis of aminoacyl-tRNAs from each amino acid and its cognate tRNA throughout an aminoacyl-adenylate intermediate –, and many protein factors that participate in the different biosynthetic phases – initiation, elongation, translocation, and termination of the polypeptide chain.

*In vitro* selection experiments have shown the ability of ribozymes to catalyze the basic steps of translation: RNA aminoacylation – including the formation of the activated aminoacyl-adenylate – and peptide bond synthesis (see Joyce, 2002, and references therein). Thus, the emergence of this metabolic process seems chemically plausible in an RNA world (Lazcano et al., 1992; Maynard Smith and Szathmáry, 1995; Brosius, 2001; Joyce, 2002; Fox and Naik, 2004; Noller, 2004).

Some models try to present the evolutionary transitions during the establishment of all the macromolecular components of the translation machinery and the origin of the genetic code in the context of the RNA world hypothesis. Thus, the key role of tRNAs for the emergence of the primitive ribosome in an RNA world has been emphasized by several authors (Weiner and Maizels, 1987; Brosius, 2001). The coding coenzyme handle (CCH) hypothesis is a testable scheme for the origin of translation and the code (Szathmáry, 1999). In short, this proposal starts with the classical notion that present day coenzymes (carrying a ribonucleotide moiety) are molecular fossils of the RNA world (White 1976, 1982). In the primitive stage of an RNA world, ribozymic activities would be supplemented with amino acids and short peptides acting as cofactors, leading to further metabolically complex stages. The CCH hypothesis suggests that the binding of cofactors to the ribozymes was non-covalent, through base-pairing of short oligonucleotides recognizing some sequence of the ribozyme active site. The progressive replacement of ribozymes by polypeptides led to the transition to an RNA–protein world with an incipient genetic code initially established by the recognition of the oligonucleotide handles (precursors of anticodons) of the coenzymes and the handle binding sites (precursors of codons) of the ribozymes (for further details see Szathmáry and Maynard Smith, 1997; Szathmáry, 1999). In summary, translation might initially have evolved to rise the functional versatilities of the RNA world (Noller, 2004) but eventually also sparked off its fall (Joyce, 2002).



#### 5.6.1.2. *Origin and evolution of the genetic code*

The origin of the genetic code – i.e. the assignment of base triplets to amino acids during protein biosynthesis – remains a mystery since its deciphering in the 1960s. One of the most impressive characteristics of the code is its universality: except for some recent evolutionary innovations, notably mitochondrial variants, all extant organisms use the same code (Santos and Tuite, 2004). This is one of the most compelling arguments favoring the existence of a universal ancestor (see part 5.7). However, the classic hypothesis by Crick (1968) suggesting that the code is a *frozen accident* (i.e. an historical accident fixed in the universal ancestor) has been challenged by models with different balance between chance and necessity during the origin and evolution of the code and, especially, by the very occurrence of code variants (Santos and Tuite, 2004).

#### 5.6.1.3. *An expanding code*

In their classical paper on the hypercycle and the origin of genetic information Eigen and Schuster (1978) suggested that the primitive code did use units of less than three bases. It follows that the number of initial codons (and coded amino acids) was less than the current 64 (for the 20 amino acids universally found in proteins). During its early evolution the code would have increased both the number of codons and of coded amino acids, and the present code would reflect the pattern of this historical expansion (for recent hypotheses see: Patel, 2005 and Wu et al., 2005).

#### 5.6.1.4. *The stereochemical hypothesis*

Pelc (1965) and Dunnill (1966) postulated the appearance of a primitive code based in the specific steric interaction between base triplets – codons or anticodons – and amino acids. Nowadays one experimental test of this model searches the synthetic RNA sequences that bind strongly an amino acid. The *in vitro* selection methods have originated several aptamers – i.e. specific RNA ligands – for some protein amino acids, like Phe, Ile, His, Leu, Gln, Arg, Trp, and Tyr, that contain the coding sequences (codon and/or anticodon) for those amino acids (Yarus et al., 2005), but this method does not work in other cases, like Val (Yarus, 1998). In general, it is assumed that stereochemical interactions played some role – more or less strong, more or less visible in present-day cells – during the origin and early evolution of the code. A critical analysis on the stereochemical hypothesis can be found in Ellington et al. (2000).

#### 5.6.1.5. *Adaptive evolution*

The degeneracy of the code (i.e. the existence of synonymous codons) and the similarity of the codons (one base difference) for physicochemically similar

amino acids suggest that some optimization process has sculpted most of the code to minimize the damage due to point mutations (Sonneborn, 1965) or mistranslations (Woese, 1965). Recent approaches to test this idea consist in the statistical comparison of the natural code with thousands of computationally generated random codes. The results suggest that nature's choice might be the best possible code (Freeland et al., 2003). Historical, stereochemical, and adaptive patterns have been combined into a coherent picture by Knight et al. (1999).

#### 5.6.1.6. *The coevolution model*

Degeneracy aside, there are other patterns observed in the table of the genetic code (for a good summary, see Figure 1 in Szathmáry, 1999). Thus, amino acids adjacent in biosynthetic pathways cluster together in the code (Dillon, 1973; Jukes, 1973; Wong, 1975). Prebiotic chemistry on early Earth did not supply all 20 current protein amino acids (see part 5.2) and most likely some of them had a biosynthetic origin. The expanding amino acid repertoire might have therefore coevolved with the code, namely new metabolic products could usurp codons previously used by their metabolic precursors (for a recent review, see Wong, 2005).

#### 5.6.1.7. *Clues from aminoacyl-tRNA synthases and tRNAs*

Aminoacyl-tRNA synthases (aaRS) are responsible of the actual decoding: in extant cells, each one of these 20 enzymes specifically recognizes each protein amino acid and their corresponding cognate tRNAs. A notable observation is the existence of a set of rules through which current aaRS recognize the tRNA molecules, mostly at the end of the acceptor stem – i.e. the end part of the tRNA that it is esterified by the amino acid, far from the anticodon stem, the part that recognizes the codon in the mRNA. Those rules – also known as the *operative code* – were established using minihelices, i.e. small fragments of RNA that mimic the aforementioned acceptor stems (Schimmel et al., 1993). On the other hand, structural studies of the active sites have revealed the existence of two classes of aaRS, each deriving from a different, non-homologous, ancient protein domain (Schimmel et al., 1993). Members of the two classes differ in the binding region on the acceptor stem of tRNAs, so that each tRNA can be potentially recognized by two binding sites, each one from one class, in a symmetric way. The specific pairings of the two aaRS classes to the different tRNAs have served to unveil a new pattern in the code. Ribas de Pouplana and Schimmel (2001) convincingly argued that the origin and coevolution of the aaRS-tRNA specific recognitions left their imprints in the present-day code. Thus, in a primitive stage each tRNA was recognized by two ancestral aaRS active sites, each from a different class, and

each amino acid of the primitive – smaller – set had several codons assigned. The incorporation of new amino acids required the duplication of the elements (aaRS and tRNAs), leading to both the redistribution of codons and aaRS classes for the new couples amino acid/tRNA to originate the extant code. Subsequently, aaRS acquired new domains to better recognize both amino acid lateral chains and cognate tRNAs.

In summary, most of our current ideas about the origin of the genetic code are refinements of the early proposals made just after the code was established more than 30 years ago. The different hypotheses are complementary rather than mutually exclusive and the processes postulated could have been simultaneous. Although we may never know exactly how and when the universal code was established, experimental testing of the biochemical plausibility of the different proposals may eventually contribute to complete a coherent evolutionary narrative with component parts from all the hypotheses.

#### 5.6.2. DATING GENETIC TAKEOVERS: HOW OLD ARE CELLULAR DNA GENOMES?

ANTONIO LAZCANO

As demonstrated by the numerous double-stranded polymeric structures, with backbones quite different from those of nucleic acids but held together by Watson–Crick base pairing, that have been synthesized in the past few years, a wide variety of informational molecules that have the potential for genetic information transfer is possible. These nucleic acid analogues provide useful laboratory models of the molecules that may have bridged the gap between the prebiotic soup and the earliest living systems. However, nowadays the only genetic polymers found in biological beings are RNA and DNA, two closely related nucleic acids. Which of them is older? Although the possibility of a simultaneous origin of DNA and RNA has been suggested (Oró and Stephen-Sherwood, 1974), many agree with the pioneering proposals made by A. N. Belozerskii and J. Brachet at the 1957 Moscow meeting of the origin of life, who suggested in an independent way that the high amounts of cellular RNA could be interpreted as evidence of a more conspicuous role during early biological evolution (cf. Dowerkin et al., 2003). This possibility is strongly supported by deoxyribonucleotide biosynthesis, a highly conserved pathway that forms the monomeric constituents of DNA via the enzymatic reduction of RNA's monomers. As reviewed elsewhere (Lazcano et al., 1988), this metabolic pathway can be interpreted as evidence that the transition from RNA to DNA genomes did not involve major metabolic changes but was facilitated by the evolutionary acquisition of one single enzymatic step (Lazcano et al., 1988).

During the past 15 years the possibility of early replicating and catalytic cell systems based on RNA and devoid of both proteins and DNA, as suggested in the late 1960's by Carl R. Woese, Francis H. Crick and Leslie E. Orgel, i.e. the so-called RNA world, has received considerable support with the demonstration of wide range of biochemical reactions catalyzed by ribozymes (Joyce, 2002). Although the nature of the predecessor(s) of the RNA world are completely unknown and can only be surmized, it is reasonable to assume that the extreme chemical instability of catalytic and replicative polyribonucleotides may have been the main limitation of such RNA world. In fact, the chemical lability of RNA implies that primordial ribozymes must have been very efficient in carrying out self-replication reactions in order to maintain an adequate inventory of molecules needed for survival. The instability of RNA may have been one of the primary reasons underlying the transition to the extant DNA/RNA/protein world where, due to the increased stability of the genetic molecules, survival would have been less dependent on polymer stability.

How did such transition take place? According to Joyce (2002), it is possible that in the RNA world ribozymes arose that could catalyze the polymerization of DNA; in this manner information stored in RNA could be transferred to the more stable DNA. On the other hand, since four of the central reactions involved in protein biosynthesis are catalyzed by ribozymes, their complementary nature suggest that they may have first appeared in an RNA world (Kumar and Yarus, 2001), i.e. double-stranded DNA genomes appeared once ribosome-catalyzed, nucleic acid-coded protein synthesis had evolved in an RNA-dominated primitive biosphere. Thus, the most likely explanation for DNA takeover could have been that because of increased stability much longer oligomers could have accumulated and this provided for an enhanced storage capacity of information that could be passed on to the next generation of living entities. Before long, RNA which once played the singular role of replication and catalysis was replaced by the more efficient and robust DNA/protein world wherein RNA was demoted to a role of messenger/transcriber of DNA stored information needed for protein biosynthesis.

When did the transition from RNA to DNA cellular genomes take place? As is often the case with metabolic innovations, such antiquity of such step cannot be documented from geological data. However, since all extant cells are endowed with DNA genomes, the most parsimonious conclusion is that this genetic polymer was already present in the last common ancestor (LCA) that existed prior to the divergence of the three primary domains, i.e., the Bacteria, Archaea, and Eucarya. As discussed elsewhere (Delaye et al., 2004), although there have been a number of suggestions that the LCA (or its equivalents) was endowed with genomes formed by small-sized RNA molecules or hybrid RNA/DNA genetic system, there are manifold indications

that double-stranded DNA genomes of monophyletic had become firmly established prior to the divergence of the three primary domains. The major arguments supporting this possibility are:

- (a) in contrast with other energetically favourable biochemical reactions (such as hydrolysis of the phosphodiester backbone, or the transfer of amino groups), the direct removal of the oxygen from the 2'-C ribonucleotide pentose ring to form the corresponding deoxy-equivalents is a thermodynamically much less-favoured reaction. This is a major constraint that strongly reduces the likelihood of multiple, independent origins of biological ribonucleotide reduction;
- (b) the demonstration of the monophyletic origin of ribonucleotide reductases (RNR), which are the enzymes involved in the biosynthesis of deoxyribonucleotides from ribonucleotide precursors, is greatly complicated by their highly divergent primary sequences and the different mechanisms by which they generate the substrate 3'-radical species required for the removal of the 2'-OH group. However, sequence analysis and biochemical characterization of RNRs from the three primary biological domains has confirmed their structural similarities, which speaks of their ultimate monophyletic origin (see Freeland et al., 1999 and references therein);
- (c) the sequence similarities shared by many ancient, large proteins found in all three domains suggest that considerable fidelity existed in the operative genetic system of their common ancestor (i.e., the LCA), but such fidelity is unlikely to be found in RNA-based genetic systems (Lazcano et al., 1992).

The nucleic acid replication machinery requires, at the very least, a set of enzymes involving a replicase, a primase, and a helicase. Quite surprisingly, when the corresponding enzymes are compared between the three primary domains, they appear to be of polyphyletic origins although there are indications of a closer phylogenetic relationship between the Eucarya and the Archaea. Given the central role that is assigned to nucleic acid replication in mainstream definitions of life, the lack of conservation and polyphyly of several of its key enzymatic components is somewhat surprising. However, there is structural evidence that some of components do have a single origin. It is reasonable to assume, for instance, that the oldest function of polymerases is the formation of the phosphodiester bond in a growing nucleic acid molecule. This reaction is catalyzed by the so-called palm domain of DNA polymerase I, and there is evidence of the ample phylogenetic distribution of this subdomain not only in the homologues of bacterial DNA polymerase I, but also among mitochondrial and viral DNA-dependent RNA polymerases (Steitz, 1999). These findings have been interpreted to imply that the palm domain is one of the oldest recognizable components of an ancestral cellular polymerase, that may have acted both as a replicase and a transcriptase during the RNA/protein

world stage (Delaye et al., 2004). The presence of homologous palm domain in DNA- and RNA polymerases suggests that once the advent of double-stranded DNA took place, relatively few mutations would have been required for the evolution of this RNA replicase into a DNA polymerase, well before the divergence of the three domains. If this hypothesis is correct, then the palm domain of extant DNA polymerases is a silent but chemically active witness of the nature of biological systems older than the double stranded DNA cellular genomes. The biosphere never loses the traces of its evolutionary past.

### 5.7. The Last Common Ancestor

DAVID MOREIRA AND PURIFICACIÓN LÓPEZ-GARCÍA

In *The origin of species*, Darwin concluded that a logical outcome of the premises of descent with modification and natural selection is ‘*that probably all the organic beings which have ever lived on this earth have descended from some one primordial form, into which life was first breathed*’ (Darwin, 1859). The strong development of biochemistry and molecular biology during the last century clearly demonstrated that all terrestrial life is based on common biochemical themes (Kornberg, 2000; Pace, 2001), lending solid ground to that idea. Since the basic structural constituents of the cell and its most fundamental metabolic reactions are shared by the three domains of life, Bacteria, Eucarya and Archaea (Woese and Fox 1977b), the hypothesis of a common ancestor from which modern organisms possessing already those traits made its way naturally.

Such hypothetical ancestor has received different names: the last common ancestor, the cenancestor (from the Greek *kainos* meaning recent and *koinos* meaning common) (Fitch and Upper, 1987) and the last universal common ancestor or LUCA (Forterre and Philippe, 1999). Although impossible to date, the cenancestor likely existed several billion years ago, predating the diversification of the three life domains. Therefore, the reconstruction of its model portrait is a difficult, speculative, task. Most authors favour the idea that the cenancestor was a single organism, an individual cell that existed at a given time and that possessed most of the features (and the genes encoding them) common to all contemporary organisms (Figure 5.10A–D). Others, on the contrary, envisage a population of cells which, as a whole, possessed all those genes, although no single individual did (Kandler, 1994; Woese, 1998; Woese, 2000). The level of gene exchange and spreading in this population should have been very high but, at some point, a particular successful combination of genes occurred in a subpopulation that became ‘isolated’ and gave rise to a whole line of descent. Kandler, for instance, proposed in his ‘pre-cell theory’ that

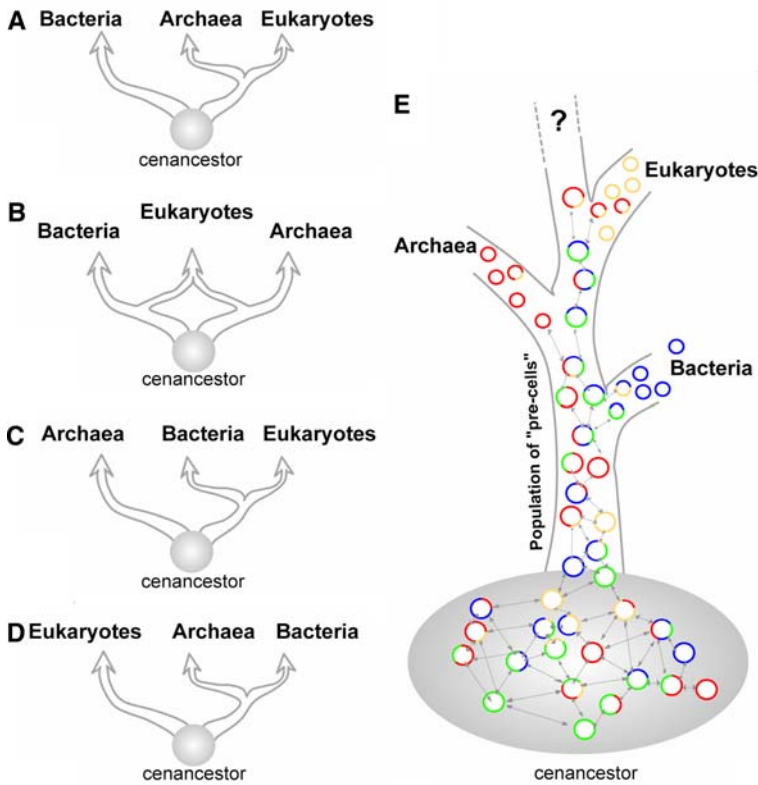


Figure 5.10. Different models of evolution of the three domains of life from a single-cell cenancestor (A–D) or a population of pre-cells (E).

bacteria, archaea, and eukaryotes emerged sequentially in this way (Figure 5.10E) (Kandler, 1994; Wächtershäuser, 2003). Despite the controversy between the ‘single-cell’ and ‘population’ hypotheses, the cenancestor is generally conceived as an already quite complex entity. This implies that life had already undergone a more or less long evolutionary pathway from the prebiotic times to the cenancestor stage. Therefore, both the origin of life and the nature of the cenancestor are different evolutionary questions. However, although rather complex, the cenancestor’s level of complexity depends on the model. For Woese, it was a relatively primitive entity, a ‘progenote’, that ‘*had not completely evolved the link between genotype and phenotype*’ (Woese and Fox, 1977a). For many other authors the progenote state occurred prior to the cenancestor, which was nearly a modern cell (Doolittle, 2000a, b).

Is it possible to build a model portrait of the cenancestor? At least, several of its fundamental characteristics can be inferred thanks to biochemistry and molecular biology comparative studies and, more recently, to the most



powerful comparative genomics and molecular phylogeny. This approach has provided strong evidence for the very ancient origin of the ribosome-based protein synthesis (translation), a well-developed transcription machinery for the synthesis of structural and messenger RNAs, and the energy-obtaining process based on the generation of a proton gradient across membranes. All these are universal features in contemporary cells and, consequently, most likely present in the cenancestor. Other cenancestor's properties are, however, much more controversial, such as the existence of a DNA-based genome or even the possession of lipid-based membranes.

### 5.7.1. PROTEIN SYNTHESIS IN THE CENANCESTOR

When all available complete genome sequences of bacterial, archaeal and eukaryotic species are compared, only ~60 genes are found to be common to all of them. This is a small number knowing that prokaryotic species contain between 500 and 10,000 genes, and eukaryotic species between 2,000 and 30,000 genes. Interestingly, the set of 'universal' genes is almost entirely integrated by genes encoding ribosomal RNA and ribosomal proteins and other proteins involved in translation (aminoacyl-tRNA synthetases and translation factors) (Koonin, 2003). These genes are most likely ancestral, strongly suggesting that the cenancestor possessed a ribosomal-based translation machinery for protein synthesis comparable to that found in modern organisms. Protein synthesis by ribosomes is, therefore, the most universally conserved process and it appears to have remained practically unchanged since the cenancestor's times.

A few genes of that universal core encode RNA polymerase subunits, responsible for the synthesis of messenger and other RNAs from their DNA templates (genes). This suggests that the cenancestor possessed at least part of the transcription machinery that is found in contemporary organisms. Nevertheless, several RNA polymerase subunits and transcription factors are not universally distributed, so that the evolutionary conservation of the transcription machinery is not as high as that of translation.

### 5.7.2. THE CENANCESTOR'S GENOME: DNA OR RNA?

DNA is the molecule where genetic information is stored in all contemporary cells. However, only three out of the ~60 universal genes are related to DNA replication and/or repair: one DNA polymerase subunit, one exonuclease and one topoisomerase (Koonin, 2003). Today, archaea and eukaryotes share many genes involved in DNA replication that are absent from bacteria which, in turn, possess apparently unrelated genes encoding proteins performing equivalent functions. Various hypotheses try to explain this paradox. One of

them postulates that the cenancestor did not possess a DNA, but an RNA, genome. DNA and its replication would have evolved twice independently, once in the bacterial line of descent and other in a lineage leading to archaea and eukaryotes (Leipe et al., 1999; Forterre, 2002). However, although few, universally conserved proteins and protein domains involved in DNA metabolism exist (Giraldo, 2003), suggesting that the cenancestor possessed indeed a DNA genome. Furthermore, due to a higher mutation rate, RNA is much more error-prone than DNA and, consequently, individual RNA molecules cannot exceed a certain size (Eigen limit) without falling into replicative catastrophe (Eigen, 1971; 2002). This size is small (~30–50 kb), so that a single RNA molecule can contain only a few dozens of genes. Since the cenancestor may have had over 600 genes (Koonin, 2003), its genome would have required many RNA molecules, entailing serious problems of stability and partition among daughter cells. Contemporary DNA and RNA viruses are good examples for these stability problems: DNA viruses can have very large genomes, up to ~1 Mbp (Raoult et al., 2004), but those of RNA viruses do not exceed ~30 kb (Domingo and Holland, 1997).

The remaining models postulate a DNA-based genome for the cenancestor. One of them hypothesises that transcription and translation were already well developed while DNA replication was still very primitive. DNA replication would have been refined as the two lines of descent leading to the bacteria and to the archaea/eukaryotes diverged (Olsen and Woese, 1997). An opposite possibility could be that the cenancestor had a very complex DNA-based metabolism, containing ancestral versions of the proteins found today in both the bacterial and the archaeal/eukaryotic replication machineries. One set of these proteins would be involved in replication, whereas the other would be specialized in DNA repair. During the separation of the two lineages, each line of descent would have retained only one of the two sets of proteins. In another hypothesis, the DNA replication machinery would be already developed in the cenancestor, but it evolved very fast in one or the two lineages descending from the cenancestor, so that the similarity between homologous genes in the two lineages are no longer recognisable (Olsen and Woese, 1997; Moreira, 2000). Yet in another model, the cenancestor would have had a DNA replication machinery, whose genes were inherited by archaea and eukaryotes, but that was replaced by viral genes in bacteria (Forterre, 1999). However, the directionality of these gene transfers is highly discussed, since present evidence demonstrates that viruses have acquired many, including replication-related, genes from their bacterial hosts (Moreira, 2000).

### 5.7.3. ENERGY AND CARBON METABOLISM IN THE CENANCESTOR

The question of how metabolism looked like in the cenancestor is rarely addressed, partly because the reconstruction of early metabolic pathways is

extremely difficult. Genes involved in energy and carbon metabolism most often display a patchy distribution in organisms of the three domains of life and usually belong to large multigenic families whose members have been recruited for different functions in various metabolic pathways. In addition, horizontal gene transfer frequently affects metabolic genes, since they may confer an immediate adaptive advantage to the new host. Hence, the reconstruction of ancestral metabolic pathways is hampered by a complex history of gene duplications and losses, differential enzyme recruitment, and horizontal gene transfer (Castresana and Moreira, 1999). Nevertheless, the universal presence of a highly conserved membrane-bound ATPase in contemporary organisms indicates that the cenancestor was able to produce energy in the form of ATP by generating a proton gradient across the cell membrane. The energy source required to generate this proton gradient was likely chemical (oxido-reduction – redox-reactions), since phototrophy (light-based) evolved later and only in the bacterial line. The type of electron donors and acceptors involved in those redox reactions, whether they were organic, inorganic or both, is not known. It is possible that the cenancestor used a variety of oxidized inorganic molecules as electron acceptors (Castresana and Moreira, 1999), and that it carried out a simple heterotrophic metabolism, at least some kind of fermentation.

A purely heterotrophic cenancestor, needing to uptake organic molecules from the environment, would be logical if life originated in a prebiotic soup (Oparin, 1938; Broda, 1970). An autotrophic cenancestor, able to synthesise organic molecules from CO or CO<sub>2</sub>, would be a logical outcome of models proposing that the first living forms were chemolithoautotrophic, deriving energy from redox reactions involving inorganic molecules such as H<sub>2</sub>S, H<sub>2</sub> and FeS (Wächtershäuser, 1988b; Russell and Hall, 1997, see also part 5.3). However, this is not incompatible with soup-based models, since autotrophy might have evolved between life emergence and the cenancestor. If the cenancestor was autotrophic, which metabolic pathway to fix carbon did it use? Today, four different pathways of autotrophic carbon fixation are known: The Calvin–Benson cycle (reductive pentose-phosphate pathway), the Arnon cycle (reductive citric acid pathway), the Wood–Ljungdahl cycle (reductive acetyl-CoA pathway), and the hydroxypropionate pathway, none of which is universal (Peretó et al., 1999). The complex Calvin–Benson cycle appeared relatively late during bacterial evolution, but any of the other, simpler, pathways might have been ancestral following different authors (Wächtershäuser, 1990; Peretó et al., 1999; Russell and Martin, 2004).

#### 5.7.4. THE MEMBRANE OF THE CENANCESTOR

All contemporary cells are surrounded by a plasma membrane that is made out of phospholipids, generally organized in bilayers (see part 5.4). However

there exist profound differences between the membrane lipids of archaea and those of bacteria and eukaryotes. In archaea, phospholipids are generally made out of isoprenoid lateral chains that are bounded by ether linkages to glycerol-1-phosphate, whereas in bacteria and eukaryotes they are made out of fatty acids bounded by ester linkages to glycerol-3-phosphate. The enzymes that synthesise the glycerol-phosphate stereoisomers are not homologous in archaea and bacteria/eukaryotes, belonging to different enzymatic families. To explain this profound difference, some authors even proposed that the cenancestor was not yet membrane-bounded and that membrane lipids together with membrane-bounded cells evolved independently to generate bacteria and archaea (Koga et al., 1998). Others proposed that the cenancestor was cellular but, instead of lipid membranes, it possessed iron monosulphide boundaries. Cells would be mineral compartments in an ever-growing hydrothermal chimney traversed by alkaline fluids, and membrane lipids would have been invented independently in bacteria and archaea (Martin and Russell, 2003). Another, less radical, option is that the cenancestor had a lipid membrane that was heterochiral, i.e. composed of a mixture of lipids built upon glycerol-1-phosphate and glycerol-3-phosphate (Wächtershäuser, 2003). The biosynthetic pathways to produce the two types of homochiral membranes would have evolved as archaea and bacteria diverged. This is supported by recent phylogenetic analyses of the enzymes involved suggesting that the cenancestor synthesized phospholipids via a non-stereospecific pathway (Peretó et al., 2004). In addition, the occurrence of universally conserved membrane-bound proteins, such as the proton-pump ATPases (Gogarten et al., 1989) and the signal recognition particle, SRP (Gribaldo and Cammarano 1998) fully supports the hypothesis of a cenancestor endowed with lipid membranes.

#### 5.7.5. OTHER UNRESOLVED QUESTIONS

Many additional questions about the nature of a hypothetical cenancestor remain open, e.g. whether the cenancestor was hyperthermophilic or not, and whether it was 'simple' or 'complex'. The hypothesis of a hyperthermophilic cenancestor arose from the discovery of hyperthermophilic bacteria and archaea growing optimally at  $>80$  °C and from the proposals of a hot, autotrophic origin of life in a hotter early Earth (Achenbach-Richter et al., 1987; Pace, 1991). The first criticisms to it derived from the fact that RNA and other biomolecules have relatively short life-times at high temperatures so that a hot origin of life would be unlikely (Lazcano and Miller, 1996). However, the origin of life and the cenancestor might have occurred in different environmental conditions (Arrhenius et al., 1999). Hence, attempts to reconcile a hyperthermophilic cenancestor with a cold origin of life propose that only hyperthermophiles could survive the late heavy meteorite bombardment

~3.9 Ga ago (Gogarten-Boekels et al., 1995; about LHB, see chapter 4.4). The most decisive argument sustaining a hyperthermophilic cenancestor was that hyperthermophilic bacteria and archaea branched at the most basal positions in phylogenetic trees (Stetter, 1996). Two major objections followed. First, computer reconstruction of ancestral rRNAs suggests that the content of guanine and cytosine in the cenancestor's rRNA was incompatible with life at  $>80\text{ }^{\circ}\text{C}$  (Galtier et al., 1999), although the analysis of the same data by other methods favours, on the contrary, a hyperthermophilic cenancestor (Di Giulio, 2000; 2003). Second, refined phylogenetic analyses of rRNAs suggest that the basal emergence of hyperthermophilic bacteria was an artefact of phylogenetic tree reconstruction, and that they adapted secondarily to hyperthermophily (Brochier and Philippe, 2002). By contrast, the archaeal ancestor was most likely hyperthermophilic (Forterre et al., 2002). Therefore, if the bacterial ancestor was also a hyperthermophile, then the cenancestor was most likely hyperthermophilic too, but if the bacterial ancestor was not hyperthermophilic, the question remains open. At any rate, all models appear compatible with the occurrence of a thermophilic ( $60\text{--}80\text{ }^{\circ}\text{C}$ ) cenancestor (López-García, 1999).

Another controversial issue concerns the level of structural complexity of the cenancestor. For most authors, it was 'structurally simple', resembling today's prokaryotes, with the genetic material directly immersed in the cytoplasm. Such an ancestor would agree with the widely accepted bacterial rooting of the tree of life, between the bacteria and the archaea/eukaryotes (Figure 5.10A) (Woese et al., 1990), but would be also compatible with two alternative tree topologies (Figure 5.10B and C). However, other authors propose that the root lies between the eukaryotes and a branch leading to the two prokaryotic domains (Figure 5.10D) (Brinkmann and Philippe, 1999; Philippe and Forterre, 1999). This rooting would still be compatible with a prokaryote-like cenancestor, but opens the possibility that the cenancestor had some features of modern eukaryotes, such as a membrane-bound nucleus and many small RNAs claimed to be relics of a hypothetical RNA world (Poole et al. 1999). Despite the position of the root is indeed an open question, models proposing a eukaryotic-like cenancestor do not explain how such a complex entity was built from the prebiotic world. In this sense, a simpler, prokaryotic-like cenancestor appears much more parsimonious in evolutionary terms.

## 5.8. The Origin of Viruses

PATRICK FORTERRE

Viruses are often ignored in evolutionary scenarios. For many biologists, viruses are not even considered as genuine living beings, since they are

absolutely dependent on cellular organisms (archaea, bacteria or eukaryotes) for their development. Nevertheless, viruses are very diverse and can be quite complex, as is revealed by recent observations including (i) the discovery of a giant virus (Mimivirus) whose genome doubles the size of certain cell genomes, (ii) the fact that, after the analysis of many viral genomes, many viral proteins have no known homologues in cells, (iii) the existence of a phylogenetic link between viruses and different evolutionary distant cellular hosts (e.g. man and bacteria) or (iv) the discovery of an unexpected morphological and functional diversity, particularly in the less-explored archaeal viruses.

### 5.8.1. VIRAL PROPERTIES

Viruses may be considered as living entities if we consider their chemical composition and their developmental cycle. Viruses are made out of the same molecules as cellular organisms: proteins, nucleic acids, and often, lipids and sugars. However, viruses exhibit a number of singularities. By contrast to cellular genomes, always made out of double-stranded DNA, viral genomes can be constituted by DNA or RNA, either single- or double-stranded. The essential difference between viruses and their cellular hosts is that viruses are unable to synthesize their own proteins and to produce their ATP. Viruses lack ribosomes as well as any form of energy and carbon metabolism. Therefore, viruses are obligatory parasites of cells (Villareal, 2005).

During their cycle, viruses exhibit two different forms. One form is stable, normally extracellular, the virion, which is unable to grow and reproduce, and the other (or others) is intracellular and can lead to the viral reproduction. Virions are constituted by a protective shell, the capsid, which encapsulates the viral genome that is generally associated with proteins forming a nucleoprotein filament. Virions can be observed only by electron microscopy and display very different morphologies (spherical, filamentous, head-and-tail, etc). Being very small, they were initially discovered by their capacity to traverse filters that retain bacteria. Nonetheless, their size can vary from a few dozens to various hundreds of nanometres, thus the largest viruses are comparable in size to the smallest known cells (Villareal, 2005). Viral genomes can also vary from a few kilobase pairs for the smallest RNA or DNA viruses to 1,200 kilobase pairs for the Mimivirus genome (Raoult et al., 2005). The simplest protein capsids are assembled following a precise geometry, but many virions have one or two additional envelopes, often from cellular origin. Virions have frequently been considered as passive forms simply allowing the transport and protection of the viral genome. However, an archaeal virus has been discovered recently, whose virion is able to change its morphology extracellularly (Haring et al., 2005).

The virion can either penetrate entirely within the cell cytoplasm or just inject its genetic material into it. These processes require more or less

sophisticated mechanisms involving proteins that recognise the host cell surface, allow the fusion of the viral envelope with that of the cell, the capture of the virus by cellular endocytose or the transport of DNA through the cell plasma membrane. Once inside, the virus deviates the cellular metabolism to replicate and transcribe its own genome, make its capsid proteins and, in the end, produce new viruses. This cycle may trigger the death of infected cells with the sudden liberation of many virions (virulence), or may induce the continuous production of a limited amount of virions while allowing cell survival. Finally, the viral genome may stay within the cell without producing new viruses (lysogenic state). During the lysogenic state, viral genomes can become integrated in the host chromosome. In the case of retroviruses (e.g. AIDS virus), this integration is achieved by retrotranscription of the viral RNA to DNA. Lysogenic viruses can be reactivated when cells undergo different stress, which allows the virus to escape and initiate a new cycle.

All cellular organisms, archaea, bacteria or eukaryotes, can be infected by an extraordinary diversity of viruses. Viruses constitute indeed a major component of the biosphere and play, and have likely played, a determinant role in the evolution of their hosts (Villareal, 2005). Therefore, it is probable that viruses have also influenced early steps in biological evolution but, when and how did viruses originated? The origin of viruses is still a mystery. Many researchers think that viruses are polyphyletic, i.e. they have multiple origins. However, viruses share a number of properties that may suggest, on the contrary, a common mechanism for their emergence.

#### 5.8.2. HYPOTHESES ON THE ORIGIN OF VIRUSES

Traditionally, three hypotheses to explain the origin of viruses have been put forward. First, viruses would be primitive entities that appeared before cells. Second, viruses would derive from ancient cells that parasitized other cells by a reduction process implying the loss of their ribosomes and energy and carbon metabolism. And third, viruses would be chromosome fragments from cells that became autonomous and began to parasitize (Balzer, 2000, Forterre, 2006). The three hypotheses have been severely criticized. The first one was refuted because, being obligatory parasites, viruses could have never been emerged before cells. The second was also criticized because there are no known intermediates between viruses and cellular parasites, since even the most reduced cellular parasites (mycoplasma within the bacteria or microsporidia within the eukaryotes) have retained their basic cellular features. The third hypothesis was favoured by most biologists, although it does not explain how RNA or DNA fragments could have escaped from cells and acquire a capsid, and how the mechanisms that allow viruses to penetrate cells were developed. In addition, in its original form, this hypothesis postulated that viruses infected bacteria (bacteriophages) and those infecting



eukaryotes had originated from bacterial and eukaryotic genomes, respectively. However, many viral proteins have no known homologues in their host genomes, which exclude, for some authors, a common origin. Furthermore, some viral proteins may have homologues in a different cellular domain to that being actually infected by the virus (Bamford, 2003).

Recent structural analyses of conserved viral proteins show that RNA polymerases from RNA bacteriophages are homologous to those from RNA viruses infecting eukaryotes, and that certain capsid proteins from some DNA bacteriophages are homologous to those from DNA viruses infecting eukaryotes and archaea. Although the possibility that viruses can “jump” and infect organisms from different domains cannot be excluded, various researchers think that viruses are very ancient, existing well before the divergence of the three cellular domains, that is, before the last universal common (or cellular) ancestor (Bamford, 2003, Forterre, 2005).

All the criticism raised against the above-mentioned hypotheses is difficult to overcome if viruses emerged in the cellular world that we know today (e.g. it would be difficult to imagine a regressive evolution from modern cells towards a viral form). However, the situation may be different if viruses originated prior to a last cellular ancestor during an RNA world. Thus, some authors that support the idea of a long acellular evolution during this RNA world period have hypothesized that viruses appeared before cells. Viruses would have been “hosted” primarily by a primitive semi-liquid soup or by mineral “cells” (Koonin and Martin, 2005). This idea is criticized by other authors favouring the idea that, on the contrary, free cells surrounded by a plasma membrane emerged very early. If this was the case, two different scenarios corresponding to the second and third hypotheses mentioned above can be envisaged: viruses emerged either by reduction of RNA cells that parasitized other RNA cells, or by the separation and independence of RNA fragments that would become autonomous and infectious. These scenarios, improbable in the modern cellular world, could be more realistic in the context of primitive RNA-based cells whose genomes were likely constituted by several linear RNA molecules (Forterre, 2006).

In any case, the first viruses may have been RNA viruses, although the question of when they emerged remains unanswered. The discovery of structural homologies between viral RNA and DNA polymerases suggests that DNA viruses evolved secondarily from certain RNA viruses. It could be even possible that DNA appeared for the first time in viruses, since a chemical modification in the viral genome could have represented an immediate selective advantage in order to escape from cellular defences designed to destroy the viral genome. This hypothesis would explain why many viruses encode their own ribonucleotide reductase (the enzyme that reduces the ribose into deoxyribose) and/or thymidylate synthase (responsi-

ble for the synthesis of thymine, the nucleotide that replaces uracile in DNA) (Forterre, 2002).

Although the mechanisms that led to the emergence of viruses remain hypothetical, viruses must have played a considerable role in biological evolution. Many viral genomes or genome fragments are present (cryptic) in contemporary cell genomes. Viruses can pick up genes from one organism and transfer them to another, thus serving as vehicles of gene transfer between cellular lines. Horizontal gene transfer is now recognized as an important motor in cellular evolution. The essential role of viruses in biological history is beginning to be fairly recognized.

## References

- Achenbach-Richter, L., Gupta, R., Stetter, K. O. and Woese, C. R.: 1987, *Syst. Appl. Microbiol.* **9**, 34–39.
- Arnez, J. G. and Moras, D.: 2003, *Trends Biochem. Sci.* **22**, 211–216.
- Arrhenius, G., Bada, J. L., Joyce, G. F., Lazcano, A., Miller, S. and Orgel, L. E.: 1999, *Science* **283**, 792.
- Balter, M.: 2000, *Science* **289**, 1866–1867.
- Bamford, D. H.: 2003, *Res. Microbiol.* **154**, 231–236.
- Ban, N., Nissen, P., Hansen, J., Moore, P. B. and Steitz, T. A.: 2000, *Science* **289**, 905–920.
- Barbier, B. and Brack, A.: 1992, *J. Am. Chem. Soc.* **114**, 3511–3515.
- Bartel, D. P. and Unrau, P. J.: 1999, *Trends Biochem. Sci.* **24**, 9–13.
- Benner, S. A.: 1993, *Science* **261**, 1402–1403.
- Benner, S. A., Ellington, A. D. and Tauer, A.: 1989, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7054–7058.
- Biron, J.-P. and Pascal, R.: 2004, *J. Am. Chem. Soc.* **126**, 9198–9199.
- Biron, J.-P., Parkes, A. L., Pascal, R. and Sutherland, J. D.: 2005, *Angew. Chem. Int. Ed.* **44**, 6731–6734.
- Bohler, C., Nielsen, P. E. and Orgel, L. E.: 1995, *Nature* **376**, 578–581.
- Borsenberger, V., Crowe, M. A., Lehbauer, J., Raftery, J., Helliwell, M., Bhutia, K., Cox, T. and Sutherland, J. D.: 2004, *Chem. Biodivers.* **1**, 203–246.
- Brasier, M. D., Green, O. R., Jephcoat, A. P., Kleppe, A. K., van Kranendonk, M., Lindsay, J. F., Steele, A. and Grassineau, N.: 2002, *Nature* **416**, 76–81.
- Brinkmann, H. and Philippe, H.: 1999, *Mol. Biol. Evol.* **16**, 817–825.
- Brochier, C. and Philippe, H.: 2002, *Nature* **417**, 244.
- Broda, E.: 1970, *Prog. Biophys. Mol. Biol.* **21**, 143–208.
- Brosius, J.: 2001, *Trends Biochem. Sci.* **26**, 653–656.
- Cairns-Smith, A. G.: 1966, *J. Theor. Biol.* **10**, 53–88.
- Cairns-Smith, A. G.: 1982, *Genetic Takeover and the Mineral Origins of Life*, Cambridge University Press, Cambridge, 477 pp.
- Castresana, J. and Moreira, D.: 1999, *J. Mol. Evol.* **49**, 453–460.
- Cavalier-Smith, T.: 2001, *J. Mol. Evol.* **53**, 555–595.
- Chen, I. and Szostak, J. W.: 2004, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7965–7970.
- Chen, I., Roberts, R. W. and Szostak, J. W.: 2004, *Science* **305**, 1474–1476.
- Cheng, C. M., Liu, X. H., Li, Y. M., Ma, Y., Tan, B., Wan, R. and Zhao, Y. F.: 2004, *Origins Life Evol. Biosphere* **34**, 455–464.

- Chyba, C.: 2005, *Science* **308**, 962–963.
- Chyba, C. and Sagan, C.: 1992, *Nature* **355**, 125–132.
- Commeyras, A., Collet, H., Boiteau, L., Taillades, J., Vandenabeele-Trambouze, O., Cottet, H., Biron, J.-P., Plasson, R., Mion, L., Lagrille, O., Martin, H., Selsis, F. and Dobrijevic, M.: 2002, *Polym. Int.* **51**, 661–665.
- Cordóva, A., Engqvist, M., Ibrahim, I., Casas, J. and Sundén, A.: 2005, *Chem. Commun.* 2047–2049.
- Corey, M. J. and Corey, E.: 1996, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11428–11434.
- Cornée A., Vergne J., Guyot F., Goffé B., Rouchy J.-M. and Maurel M.-C.: 2004, in G. Camoin and P. Gautret (eds.), *Microbialites and Microbial Communities in Sedimentary Systems*, Publication Spéciale ASF, Paris, 31 p.
- Crick, F. H. C.: 1968, *J. Mol. Biol.* **38**, 367–379.
- Cronin J. R. and Chang S. 1993, in J. M. Greenberg, C. X. Mendoza-Gomez and V. Pirronello (eds.), *The Chemistry of Life's Origins*, Kluwer Dordrecht pp. 209–258.
- Darwin, C.: 1859, *The Origin of Species by Means of Natural Selection*, J. Murray, London.
- Deamer, D. W.: 1986, *Origins Life Evol. Biosphere* **17**, 3–25.
- Deamer, D. W.: 1997, *Microbiol. Mol. Biol. Rev.* **61**, 239–261.
- Deamer, D., Dworkin, J. P., Sandford, S. A., Bernstein, M. P. and Allamandola, L. J.: 2002, *Astrobiology* **2**, 371–381.
- Décout, J.-L., Vergne, J. and Maurel, M.-C.: 1995, *Macromol. Chem. Phys.* **196**, 2615–2624.
- Delaye L., Becerra A. and Lazcano A.: 2004, in L. Ribas de Pouplana (ed.), *The Genetic Code and the Origin of Life*. Landes Bioscience, Georgetown, pp. 34–47.
- Di Giulio, M.: 2000, *J. Theor. Biol.* **203**, 203–213.
- Di Giulio, M.: 2003, *J. Theor. Biol.* **224**, 277–283.
- Di Giulio, M.: 2005, *Gene* **346**, 7–12.
- Dillon, L. S.: 1973, *Bot. Rev.* **39**, 301–345.
- Domingo, E. and Holland, J. J.: 1997, *Annu. Rev. Microbiol.* **51**, 151–178.
- Doolittle, R. F.: 2000a, *Res. Microbiol.* **151**, 85–89.
- Doolittle, W. F.: 2000b, *Curr. Opin. Struct. Biol.* **10**, 355–358.
- Dunnill, P.: 1966, *Nature* **210**, 1267–1268.
- Dworkin, J. P., Lazcano, A. and Miller, S. L.: 2003, *J. Theor. Biol.* **222**, 127–134.
- Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S. M., Driver, D. A., Berg, R. H., Kim, S. K., Norden, B. and Nielsen, P. E.: 1993, *Nature* **365**, 566–568.
- Eigen, M.: 1971, *Naturwissenschaften* **58**, 465–523.
- Eigen, M.: 2002, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13374–13376.
- Eigen, M. and Schuster, P.: 1978, *Naturwissenschaften* **65**, 341–369.
- El Amri, C., Baron, M.-H. and Maurel, M.-C.: 2003, *Spectrochim. Acta Part A* **59**, 2645–2654.
- El Amri, C., Baron, M.-H. and Maurel, M.-C.: 2004, *J. Raman Spectrosc.* **35**, 170–177.
- El Amri, C., Maurel, M.-C., Sagon, G. and Baron, M.-H.: 2005, *Spectrochim. Acta Part A* **61**, 2049–2056.
- Ellington, A. D., Khrapov, M. and Shaw, C. A.: 2000, *RNA* **6**, 485–498.
- Eschenmoser, A.: 1999, *Science* **284**, 2118–2124.
- Ferris, J. P.: 1987, *Cold Spring Harbor Symp. Quant. Biol.* **LII**, 29–39.
- Ferris, J. P. and Ertem, G.: 1992, *Science* **257**, 1387–1389.
- Ferris, J. P., Hill, A. R., Liu, R. and Orgel, L. E.: 1996, *Nature* **381**, 59–61.
- Fersht, A. R.: 1999, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, Freeman, New York.
- Fitch, W. M. and Upper, K.: 1987, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 759–767.
- Forterre, P.: 1999, *Mol. Microbiol.* **33**, 457–465.

- Forterre, P.: 2002, *Curr. Opin. Microbiol.* **5**, 525–532.
- Forterre, P.: 2005, *Biochimie* **87**, 793–803.
- Forterre P.: 2006, *Virus Res.*, 5–16.
- Forterre, P. and Philippe, H.: 1999, *Biol. Bull.* **196**, 373–375.
- Forterre, P., Brochier, C. and Philippe, H.: 2002, *Theor. Popul. Biol.* **61**, 409–422.
- Fox G. E. and Naik A. K. 2004, in Ll. Ribas de Pouplana (ed.), *The Genetic Code and the Origin of Life*, Landes Bioscience, Georgetown, pp. 92–106.
- Freeland, S. J., Knight, R. D. and Landweber, L. F.: 1999, *Science* **286**, 690–692.
- Freeland, S. J., Wu, T. and Keulman, N.: 2003, *Origins Life Evol. Biosphere* **33**, 457–477.
- Fuller, W. D., Sanchez, R. A. and Orgel, L. E.: 1972, *J. Mol. Biol.* **67**, 25–33.
- Galtier, N., Tourasse, N. and Gouy, M.: 1999, *Science* **283**, 220–221.
- Gesteland, R. F., Cech, T. R. and Atkins, J. F., (eds.): 1999, *The RNA World* (2nd ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.
- Gilbert, W.: 1986, *Nature* **319**, 618.
- Giraldo, R.: 2003, *FEMS Microbiol. Rev.* **26**, 533–554.
- Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E. J., Bowman, B. J., Manolson, M. F., Poole, R. J., Date, T., Oshima, T., Konishi, J., Denda, K. and Yoshida, M.: 1989, *Proc. Nat. Acad. Sci. U.S.A.* **86**, 6661–6665.
- Gogarten-Boekels, M., Hilario, E. and Gogarten, J. P.: 1995, *Origins Life Evol. Biosphere* **25**, 251–264.
- Gribaldo, S. and Cammarano, P.: 1998, *J. Mol. Evol.* **47**, 508–516.
- Hanczyc, M. M. and Szostak, J. W.: 2004, *Curr. Opin. Chem. Biol.* **8**, 660–664.
- Hanczyc, M. M., Fujikawa, S. M. and Szostak, J. W.: 2003, *Science* **302**, 618–622.
- Haring, M., Vestergaard, G., Rachel, R., Chen, L., Garrett, R. A. and Prangishvili, D.: 2005, *Nature* **436**, 1101–1102.
- Harold, F. M.: 2001, *The Way of the Cell. Molecules, Organisms and the Order of Life*, Oxford University Press, Oxford.
- Healy, V. L., Mullins, L. S., Li, X., Hall, S. E., Raushel, F. M. and Walsh, C. T.: 2000, *Chem. Biol.* **7**, 505–514.
- Hill, A. R., Orgel, L. E. and Wu, T.: 1993, *Orig. Life Evol. Biosphere* **23**, 285–290.
- Hoang, L., Fredrick, K. and Noller, H. F.: 2004, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12439–12443.
- Holm, N. G. and Andersson, E. M.: 1998, in A. Brack (ed.), *The Molecular Origins of Life Assembling pieces of the Puzzle*, Cambridge University Press, Cambridge, pp. 86–99.
- Huber, C., Eisenreich, W., Hecht, S. and Wächtershäuser, G.: 2003, *Science* **301**, 938–940.
- Inoue, T. and Orgel, L. E.: 1983, *Science* **219**, 859–862.
- Joyce, G. F.: 1989, *Nature* **338**, 217–224.
- Joyce, G. F.: 2002, *Nature* **418**, 214–221.
- Joyce, G. F. and Orgel, L. E.: 1986, *J. Mol. Biol.* **188**, 433–441.
- Joyce, G. F. and Orgel, L. E.: 1999, in R. F. Gesteland, T. R. Cech and J. F. Atkins (eds.), *The RNA World*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY, pp. 49–77.
- Joyce, G. F., Schwartz, A. W., Miller, S. L. and Orgel, L. E.: 1987, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4398–4402.
- Jukes, T. H.: 1973, *Biochem. Biophys. Res. Commun.* **53**, 709–714.
- Kandler O.: 1994, in: S. Bengston (ed.), *Early Life on Earth*, Columbia University Press, pp. 152–160.
- Kasting, J. F.: 1993, *Science* **259**, 920–926.
- Knight, R. D., Freeland, S. J. and Landweber, L. F.: 1999, *Trends Biochem. Sci.* **24**, 241–247.
- Koga, Y., Kyuragi, T., Nishihara, M. and Sone, N.: 1998, *J. Mol. Evol.* **46**, 54–63.

- Koonin, E. V.: 2003, *Nat. Rev. Microbiol.* **1**, 127–136.
- Koonin, E. V. and Martin, W.: 2005, *Trends Genet.* **21**, 647–654.
- Koppitz, M., Nielsen, P. E. and Orgel, L. E.: 1998, *J. Am. Chem. Soc.* **120**, 4563–4569.
- Kornberg, A.: 2000, *J. Bacteriol.* **182**, 3613–3618.
- Kuhn, H. and Kuhn, C.: 2003, *Angew. Chem. Int. Ed.* **42**, 262–268.
- Kühne, H. and Joyce, G. F.: 2003, *J. Mol. Evol.* **57**, 292–298.
- Kumar, R. K. and Yarus, M.: 2001, *Biochemistry* **40**, 6998–7004.
- Lahav, N., White, D. and Chang, S.: 1978, *Science* **201**, 67–69.
- Lazcano A. 2001, in D. E. G. Briggs and P. R. Crowther (eds.), *Paleobiology II*, Blackwell Science Oxford, ch. 1.
- Lazcano, A. and Miller, S. L.: 1996, *Cell* **85**, 793–798.
- Lazcano, A., Guerrero, R., Margulis, L. and Oró, J.: 1988, *J. Mol. Evol.* **27**, 283–290.
- Lazcano, A., Fox, G. E. and Oró, J.: 1992, in R. P. Mortlock (ed.), *The Evolution of Metabolic Function*, CRC Press, Boca Raton FL, pp. 237–295.
- Leipe, D. D., Aravind, L. and Koonin, E. V.: 1999, *Nucleic Acids Res.* **27**, 3389–3401.
- Leman, L., Orgel, L. and Ghadiri, M. R.: 2004, *Science* **305**, 283–286.
- López-García, P.: 1999, *J. Mol. Evol.* **49**, 439–452.
- Luisi, P. L.: 1998, *Origins Life Evol. Biosphere* **28**, 613–622.
- Luisi, P. L., Walde, P. and Oberholzer, T.: 1994, *Ber. Bunsenges. Phys. Chem.* **98**, 1160–1165.
- Luisi, P. L., Walde, P. and Oberholzer, T.: 1999, *Curr. Opin. Colloid Interface Sci.* **4**, 33–39.
- Madigan, M. T., Martinko, J. M. and Parker, J.: 2003, *Brock Biology of Microorganisms*, Prentice Hall Upper Saddle River, NJ, USA.
- Marahiel, M. A., Stachelhaus, T. and Mootz, H. D.: 1997, *Chem. Rev.* **97**, 2651–2673.
- Martin, W. and Russell, M. J.: 2003, *Philos. Trans. R. Soc. Lond., B Biol. Sci.* **358**, 59–83.
- Maurel, M.-C.: 1992, *J. Evol. Biol.* **2**, 173–188.
- Maurel, M.-C. and Décout, J.-L.: 1999, *Tetrahedron* **55**, 3141–3182.
- Maurel, M.-C. and Ninio, J.: 1987, *Biochimie* **69**, 551–553.
- Maurel, M. C. and Orgel, L.: 2000, *Origins Life Evol. Biosphere* **30**, 423–430.
- Maurel, M.-C. and Zaccai, G.: 2001, *BioEssays* **23**, 977–978.
- Maynard Smith, J. and Szathmáry, E.: 1995, *The Major Transitions in Evolution*, Freeman, New York.
- McGinness, K. E., Wright, M. C. and Joyce, G. F.: 2002, *Chem. Biol.* **9**, 585–596.
- Meli, M., Vergne, J., Décout, J.-L. and Maurel, M.-C.: 2002, *J. Biol. Chem.* **277**, 2104–2111.
- Meli, M., Vergne, J. and Maurel, M.-C.: 2003, *J. Biol. Chem.* **278**, 9835–9842.
- Miller, S. L.: 1953, *Science* **117**, 528–529.
- Miller, S. L.: 1998, in A. Brack (ed.), *The Molecular Origins of Life. Assembling pieces of the Puzzle*, Cambridge University Press, Cambridge, pp. 59–85.
- Monnard, P. A. and Deamer, D. W.: 2002, *Anat. Rec.* **268**, 196–207.
- Moreira, D.: 2000, *Mol. Microbiol.* **35**, 1–5.
- Morowitz, H. J., Heinz, B. and Deamer, D. W.: 1988, *Origins Life Evol. Biosphere* **18**, 281–287.
- Morowitz, H. J., Deamer, D. W. and Smith, T.: 1991, *J. Mol. Evol.* **33**, 207–208.
- Mullie, F. and Reisse, J.: 1987, *Top. Curr. Chem.* **139**, 83–117.
- Nealson, K. H.: 2005, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3889–3890.
- Nelson, K. E., Levy, M. and Miller, S. L.: 2000, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3868–3871.
- Neunlist, S., Bisseret, P. and Rohmer, M.: 1987, *Eur. J. Biochem.* **87**, 245–252.
- Nicolis, G. and Prigogine, I.: 1977, *Self-Organization in Nonequilibrium Systems*, Wiley, New York.
- Nielsen, P. E.: 1999, *Acc. Chem. Res.* **32**, 624–630.
- Nissen, P., Hansen, J., Ban, N., Moore, P. B. and Steitz, T. A.: 2000, *Science* **289**, 920–930.

- Noller, H. F.: 2004, *RNA* **10**, 1833–1837.
- Olsen, G. J. and Woese, C. R.: 1997, *Cell* **89**, 991–994.
- Oparin, A. I.: 1938, *The Origin of Life*, Mac Millan, New York.
- Oparin, A. I., Orlovskii, A. F., Bukhlaeva, V. and Gladilin, K. L.: 1976, *Dokl. Akad. Nauk SSSR* **226**, 972–974.
- Orgel, L. E.: 1989, in M. Grunberg-Manago et al. (eds.), *Evolutionary Tinkering in Gene Expression*, Plenum Press, London, pp. 215–224.
- Orgel, L. E.: 1992, *Nature* **358**, 203–209.
- Orgel, L. E.: 2004, *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123.
- Oró, J.: 1960, *Biochem. Biophys. Res. Commun.* **2**, 407–412.
- Oró, J. and Stephen-Sherwood, E.: 1974, in J. Oró, S. L. Miller, C. Ponnampereuma and R. S. Young (eds.), *Cosmochemical Evolution and the Origins of Life*, Reidel, Dordrecht, pp. 159–172.
- Ourisson, G. and Nakatani, Y.: 1994, *Chem. Biol.* **1**, 11–23.
- Pace, N. R.: 1991, *Cell* **65**, 531–533.
- Pace, N. R.: 2001, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 805–808.
- Paecht-Horowitz, M., Berger, J. and Katchalsky, A.: 1970, *Nature* **7**, 847–850.
- Pascal, R., Boiteau, L. and Commeyras, A.: 2005, *Top. Curr. Chem.* **259**, 69–122.
- Patel, A.: 2005, *J. Theor. Biol.* **233**, 527–532.
- Pelc, S. R.: 1965, *Nature* **207**, 597–599.
- Peretó, J.: 2005, *Int. Microbiol.* **8**, 23–31.
- Peretó, J. G., Velasco, A. M., Becerra, A. and Lazcano, A.: 1999, *Int. Microbiol.* **2**, 3–10.
- Peretó, J., Lopez-Garcia, P. and Moreira, D.: 2004, *Trends Biochem. Sci.* **29**, 469–477.
- Philippe, H. and Forterre, P.: 1999, *J. Mol. Evol.* **49**, 509–523.
- Pizzarello, S. and Weber, A. L.: 2004, *Science* **303**, 1151–1151.
- Plankensteiner, K., Reiner, H. and Rode, B. M.: 2005, *Curr. Org. Chem.* **9**, 1107–1114.
- Poole, A., Jeffares, D. and Penny, D.: 1999, *Bioessays* **21**, 880–889.
- Prieur, D.: 2005, in T. Tokano (ed.), *Water on Mars and Life*, Springer, Berlin, pp. 299–324.
- Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M. and Claverie, J. M.: 2004, *Science* **306**, 1344–1350.
- Reader, J. S. and Joyce, G. F.: 2002, *Nature* **420**, 841–844.
- Reusch, R. N.: 2000, *Biochemistry (Moscow)* **65**, 280–295.
- Ribas de Pouplana, Ll. and Schimmel, P.: 2001, *Trends Biochem. Sci.* **26**, 591–596.
- Ricard, J., Vergne, J., Décout, J.-L. and Maurel, M.-C.: 1996, *J. Mol. Evol.* **43**, 315–325.
- Ricardo, A., Carrigan, M. A., Olcott, A. N. and Benner, S. A.: 2004, *Science* **303**, 196.
- Rode, B. M., Son, H. L., Suwannachot, Y. and Bujdak, J.: 1999, *Origins Life Evol. Biosphere* **29**, 273–286.
- Rohlfing, D. L.: 1976, *Science* **193**, 68–70.
- Russell, M. and Hall, A. J.: 1997, *J. Geol. Soc. Lond.* **154**, 377–402.
- Russell, M. J. and Martin, W.: 2004, *Trends Biochem. Sci.* **29**, 358–363.
- Sacerdote, M. G. and Szostak, J. W.: 2005, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6004–6008.
- Santos M. A. S. and Tuite M. F. 2004, in Ll. Ribas de Pouplana (ed.), *The Genetic Code and the Origin of Life*, Landes Bioscience, Georgetown, pp. 183–200.
- Schimmel, P., Giege, R., Moras, D. and Yokoyama, S.: 1993, *Proc. Natl. Acad. Sci. U. S. A.* **90**, 8763–8768.
- Schmidt, J. G., Nielsen, P. E. and Orgel, L. E.: 1997, *Nucleic Acids Res.* **25**, 4797–4802.
- Schöning, K.-U., Scholz, P., Guntha, S., Wu, X., Krishnamurthy, R. and Eschenmoser, A.: 2000, *Science* **290**, 1347–1351.
- Schopf, J. W.: 1993, *Science* **260**, 640–646.
- Schulze-Makuch, D. and Irwin, L. N: 2002, *Astrobiology* **2**, 105–121.



- Schwartzman, D. W. and Lineweaver, C. H.: 2004, *Biochem. Soc. Trans.* **32**, 168–171.
- Segré, D., Ben-Eli, D., Deamer, D. W. and Lancet, D.: 2001, *Origins Life Evol. Biosphere* **31**, 119–145.
- Sievers, A., Beringer, M., Rodnina, M. V. and Wolfenden, R.: 2004, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7897–7901.
- Sonneborn T. M. 1965, in V. Bryson and H. J. Vogel (eds.), *Evolving Genes and Proteins*, Academic Press, New York, pp. 379–397.
- Spear, J. R., Walker, J. J., McCollom, T. M. and Pace, N. R.: 2005, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2555–2560.
- Steitz, T. A.: 1999, *J. Biol. Chem.* **274**, 17395–17398.
- Stetter, K. O.: 1996, *FEMS Microbiol. Rev.* **18**, 149–158.
- Stoks, P. G. and Schwartz, A. W.: 1982, *Geochim. Cosmochim. Acta* **46**, 309–315.
- Sutherland, J. D. and Blackburn, J. M.: 1997, *Chem. Biol.* **4**, 481–488.
- Sutherland, J. D. and Whitfield, J. N.: 1997, *Tetrahedron* **53**, 11493–11527.
- Szathmáry, E.: 1999, *Trends Genet.* **15**, 223–229.
- Szathmáry, E. and Maynard Smith, J.: 1997, *J. Theor. Biol.* **187**, 555–571.
- Szostak, J. W., Bartel, D. P. and Luisi, P. L.: 2001, *Nature* **409**, 387–390.
- Taillades, J., Collet, H., Garrel, L., Beuzelin, I., Boiteau, L., Choukroun, H. and Commeyras, A.: 1999, *J. Mol. Evol.* **48**, 638–645.
- Tehei, M., Franzetti, B., Maurel, M.-C., Vergne, J., Hountondji, C. and Zaccai, G.: 2002, *Extremophiles* **6**, 427–430.
- Tian, F., Toon, O. B., Pavlov, A. A. and De Sterck, H.: 2005, *Science* **308**, 1014–1017.
- Tobé, S., Heams, T., Vergne, J., Hervé, G. and Maurel, M.-C.: 2005, *Nucleic Acids Res.* **33**, 2557–2564.
- Varela, F. G., Maturana, H. R. and Uribe, R.: 1974, *Biosystems* **5**, 187–196.
- Vergne J. and Maurel M.-C. 2002, *Origins Life Evol. Biosphere* (abstract) **32**(5–6), 538.
- Vergne J., Cognet J. A. H., Szathmáry E., and Maurel M.-C. 2006, *Gene*, **371**, 182–193.
- Villarreal L. P. 2005, in L. P. Villarreal (ed.), *Viruses and the Evolution of Life*, ASM Press, Washington.
- Wächtershäuser, G.: 1988a, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1134–1135.
- Wächtershäuser, G.: 1988b, *Microbiol. Rev.* **52**, 452–484.
- Wächtershäuser, G.: 1990, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 200–204.
- Wächtershäuser, G.: 2003, *Mol. Microbiol.* **47**, 13–22.
- Weber, A. L.: 2001, *Origins Life Evol. Biosphere* **31**, 71–86.
- Weber, A. L.: 2002, *Origins Life Evol. Biosphere* **32**, 333–357.
- Weiner, A. M. and Maizels, N.: 1987, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7383–7387.
- Westall, F., De Wit, M. J., Dann, J., Van der Gaast, S., De Ronde, C. and Gerneke, D.: 2001, *Precambrian Res.* **106**, 93–116.
- White, H. B.: 1976, *J. Mol. Evol.* **7**, 101–104.
- White H. B.: 1982, in J. Everse, B. Anderson and K. You (eds.), *The Pyridine Nucleotide Coenzymes*, Academic Press, New York, pp. 1–17.
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M. Jr, Morgan-Warren, R. J., Carter, A. P., Vornhein, C., Hartsch, T. and Ramakrishnan, V.: 2000, *Nature* **407**, 327–339.
- Woese, C. R.: 1965, *Proc. Natl. Acad. Sci. U.S.A.* **54**, 1546–1552.
- Woese, C.: 1998, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6854–6859.
- Woese, C. R.: 2000, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8392–8396.
- Woese, C. R. and Fox, G. E.: 1977a, *J. Mol. Evol.* **10**, 1–6.
- Woese, C. R. and Fox, G. E.: 1977b, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088–5090.
- Woese, C. R., Kandler, O. and Wheelis, M. L.: 1990, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576–4579.



- Wong, J. T.: 1975, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1909–1912.
- Wong, J. T.: 2005, *BioEssays* **27**, 416–425.
- Wu, H. L., Bagby, S. and van Elsen, D.: 2005, *J. Mol. Evol.* **61**, 54–64.
- Yarus, M.: 1998, *J. Mol. Evol.* **47**, 109–117.
- Yarus, M., Caporaso, J. G. and Knight, R.: 2005, *Annu. Rev. Biochem.* **74**, 179–198.
- Yusupov, M., Yusupova, G., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. and Noller, H. F.: 2001, *Science* **292**, 883–896.
- Zhang, B. and Cech, T. R.: 1997, *Nature* **390**, 96–100.