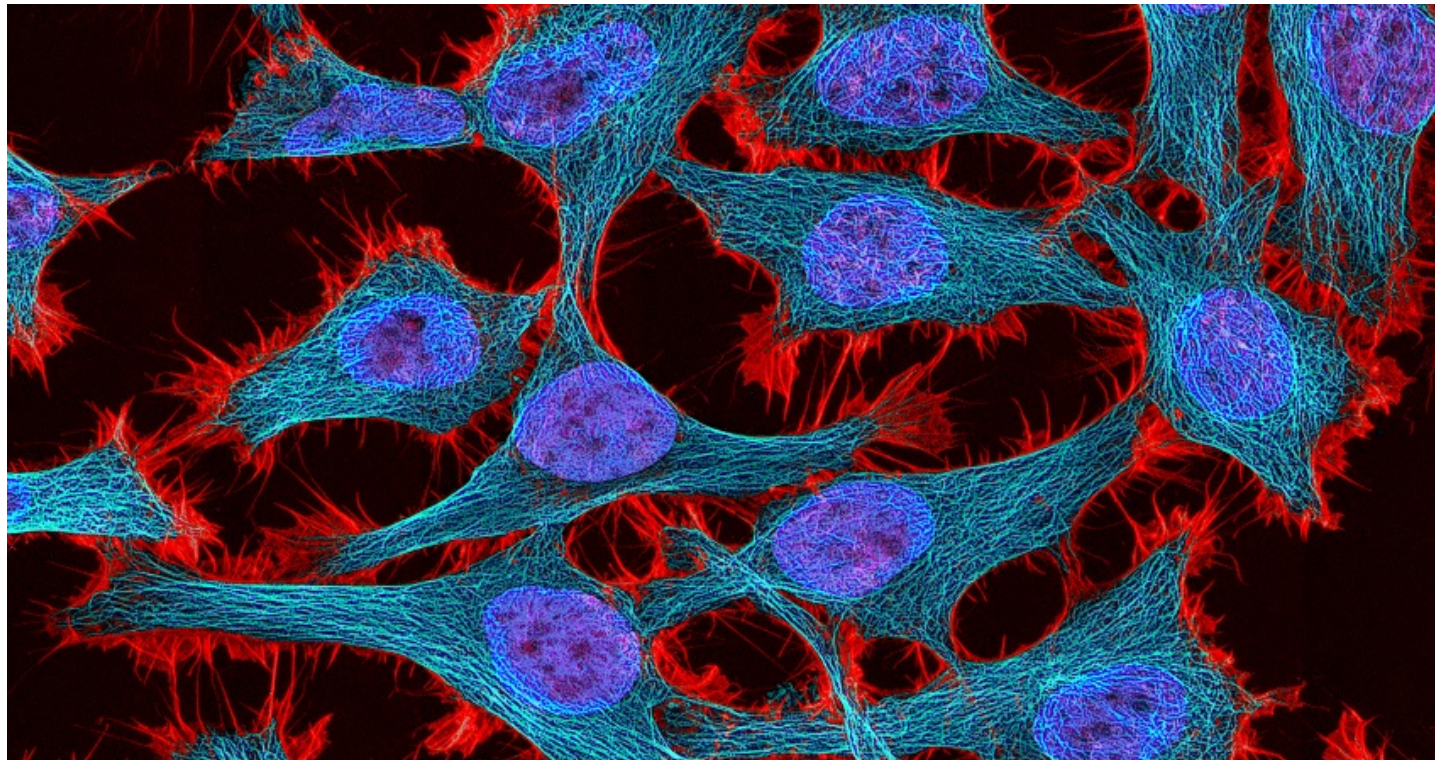


Origin of the first cells: the engineer's point of view

Auteur :

DANCHIN Antoine, Président-fondateur et directeur scientifique, Amabiotics SA ; Membre de l'Académie des Sciences

01-04-2019



In a realistic scenario of the origin of life, a primeval metabolism begins in water at the surface of minerals. These promote the polymerization of some of the elementary building blocks, including RNAs. An RNA-metabolism world develops and leads to the replication of these RNAs, eventually forming an RNA-genome world which then uses RNA molecules as information templates (primitive genes) rather than as direct substrates for metabolism. The next step is based on the invention of molecules that separate an inner medium from the outside, membrane lipids. This leads to the formation of the first cells, bringing together two compartments where metabolism and genome are separated: respectively the cytoplasm and the nucleus. A variant of RNA molecules, the celebrated DNA will allow the grouping of genes into chromosomes. Membrane lipids also allow phagocytosis and this will be the origin of the eukaryotic cell, from which plants and animals derive.

1. At the origin of life: distinguishing reproduction from replication

Presented as science-based, many scenarios propose to solve the riddle of the **origin of life**. Yet, in fact, they just reflect the *opinion* of their authors (see [The origin of life as seen by a geologist who loves astronomy](#) & [Once upon a time when life appeared....](#)). The present scenario is no exception. Before going any further, we must remove a classic pitfall, and decide whether we limit our scenarios to Earth, or whether we seek origins elsewhere in the Universe. A common way out is to propose that extraterrestrial life be the source of our terrestrial life. However, it is a way of eluding the question by moving the problem so as to make the question even more subject to pure fantasy. We will therefore follow **Ockham's razor** [1] by looking whether life is possible based on what we know about the atmosphere and the past aquatic environments of our planet. We will also hypothesize that living organisms are not palimpsests {ind-text} Manuscript consisting of a scroll already used, whose inscriptions have been removed so that they can be written on again. {end-tooltip} that have erased all memory of their past, but that they still contain today archives from the past. Also, we will restrict our quest to the atom of life, the **cell**, leaving aside multicellular

organisms.

In a little-known book, *Origins of Life*, the physicist Freeman Dyson [2] states that the information associated with life spreads along two very different axes:

1. the **reproduction** (production of a similar copy) of a metabolism (All the biochemical reactions that take place within an organism, organ or cell to enable the organism to maintain itself alive, reproduce, develop and respond to its environment. (flow of chemical transformations of carbon-based molecules);
2. the **replication of a program** (production of an identical copy, at least over a fairly long period of time).

Better still, he also demonstrates that in a realistic scenario of the origins of life, the **reproduction** of primordial processes must **predate** the emergence of a **replicative process** and then **associate** them into a coherent whole. In a chemical world such as the Earth's surface, this implies both the reproduction of chemical flows and the emergence, from this metabolism, of entities that can replicate themselves. In short, it takes at least *two* chemically distinct origins to explain the origin of life.

2. The use of genomes for a scenario of the origin of first cells

This functional scenario is very abstract, and we need to embody it into the physico-chemical reality of the Earth. Analysis of the cell shows that, whatever its origin, it is always made up of two classes of molecules, all made up of a limited number of atoms (carbon, nitrogen, oxygen, hydrogen, phosphorus and sulphur). Figure 1 represents some of the chemical elements available on Earth, a limited number of which are found in the molecules of life.

Figure 1. The atoms of life. In dark orange, the four atoms of the protein skeleton. In orange the other atoms essential to life. In light orange the atoms present in a number of organisms. Iron is almost universally present (it is missing in some bacterial species). Most of the other elements on Earth do not play an essential role in the molecules of life.

There are molecules with a few atoms ("small" molecules, metabolites) and macromolecules, made of millions, even billions of atoms. **Two classes** are to be retained here, **proteins**, made up of sequences of twenty basic units, amino acids, and **nucleic acids**, made up of chains of four distinct units, nucleotides. The sequence of macromolecules can be deciphered as a text written with an alphabet of twenty letters for proteins, and four for nucleic acids. The current intermediate metabolism (responsible for the synthesis of small molecules) thus produces the "building blocks" familiar to journalistic vocabulary: amino acids from proteins, and nucleotides involved in the synthesis of nucleic acids: ribonucleic acid (RNA) Ribonucleic acid, a macromolecule consisting of a sequence of ribonucleotides (adenine, cytosine, guanine, uracil) and performing many functions within the cell. (), and deoxyribonucleic acid (ADN) Deoxyribonucleic acid, a macromolecule that contains the genetic information of a living being. Made of two antiparallel strands wrapped around each other to form a double helix. Consists of nucleotide monomers formed of a nitrogenous base (adenine, cytosine, guanine or thymine) linked to deoxyribose, itself linked to a phosphate group. (), with nucleotide variants not to be discussed here. To do this, it uses several coenzymes (Molecules as cofactors in certain reactions catalyzed by enzymes with which they are structurally linked within a stable complex. () essential for catalysis (Action of an element that accelerates or slows a chemical reaction. () made by a huge family of proteins, enzymes. Finally, it is common sense that the world of small molecules must have preceded that of macromolecules.

Any scenario of life's beginnings should therefore provide an explanation for the occurrence of these metabolites. However, the vast majority of scenarios do not include most of them: despite the fact that carbon chemistry is commonplace in the universe, many authors are ecstatic about the presence of a few amino acids (not all of them, far from it) in various environments (very often very far from the Earth). Strangely enough, no one investigates whether they are accompanied by related molecules, foreign to life (that must therefore be kept aside like poisons are), and if they are all there. As for the origin of coenzymes, and for a long time that of nucleotides, no one really questions their origin. Finally, lipids (Hydrophobic or amphipathic molecules (having a hydrophobic part and a hydrophilic part), characterized by their insolubility in water and their solubility in non-polar organic solvents, () -molecules displaying an original behaviour towards water (hydrophilic at one end,

hydrophobic at the other end)- are needed to form the cell's envelope. The origin of these molecules is almost never contemplated. Yet, we must think that, from a limited number of basic components (including coenzymes and lipids), a primitive metabolism developed and reproduced into the ancestor of intermediate metabolism, until metabolic products (probably polymers, ancestors of proteins and nucleic acids) discovered the way to replicate within a spatial frame surrounded by a lipid membrane. The current genomes, which have been the carriers of the memory of life since its origins, provide us with the necessary ideas allowing us to understand what may have happened during these early times.

An ever-increasing number of genome sequences keeps accumulating. Comparing them with one another in the way Champollion used Rosetta stone to understand hieroglyphics allows us to recognize what is common and probably ancient in extant genomes. Unfortunately, to make things difficult, evolution tends to preserve functions, but not structures. Several objects can have the same role (we eat with a fork or with chopsticks). Also, as new genomes are sequenced, the number of genes that are deemed essential to life keeps decreasing. It has now shrunk to none! Fortunately, the comparative approach is not a total failure because some genes tend to remain present in many genomes if not all of them. These “persistent” genes encode preserved proteins because they effectively perform the most central functions. What can we say about those? In bacterial genomes, computer-based observations show that they may be grouped together according to three networks of mutual attraction.

The core network is a highly connected network, clustering the genes that manage the expression of the genetic message from its program. As cases in point, we find there the genes that drive the construction of the ribosome, the universal reading head of the RNAs carrying this message. A second, less connected network (of older origin) focuses on RNA metabolism. It defines the enzymes that nowadays couple the translation of messenger RNA (mRNA) into proteins by loading each of the twenty amino acids on a cognate RNA (transfer RNA, tRNA). In addition, connected to some of these genes we find those that organize cell division. Finally, a third set of genes that are poorly connected to one another encodes the processes allowing synthesis of the cell's central metabolites, lipids, nucleotides and amino acids, as well as the catalytic core of enzymatic proteins, coenzymes. There are also proteins whose function requires the presence of iron-sulphur centres, similar to the basic constituent of a common mineral, iron pyrite. Finally, the genes that allow the synthesis of the lipid bilayer of the cell membrane are also found in this third network.

3. From minerals to RNAs

This organization calls to mind a realistic scenario of the origin of life, based on the critical need for **compartmentalization**, as opposed to any scenario involving a prebiotic “soup”.

3.1. A dawn of stones: selection and concentration of minerals on their surface

In the first step, the mineral surface selects the reactive compounds that form the primary metabolism. The extant metabolism fully justifies this view: the majority of central metabolites are made up of negatively charged molecules (phosphates and carboxylates), whereas, very often, the chemical groups responsible for their electrical charge have no role in their function! It is therefore natural to think that these groups are the signature of a previous role, when they have been used to sort and locally concentrate molecules on the surface of minerals, allowing them to react with each other. What is more, surface metabolism creates a driving force for the construction of macromolecules through the elimination of water molecules, as observed in the formation of peptide (protein-making) and phosphodiester (ribonucleic acid-making) bonds. This tendency to polymerization results from the entropy increase associated to water escaping in the environment, in accordance with the second principle of thermodynamics. Water, for the same reason (via an increase, not a decrease in entropy, contrary to an unfortunately widespread misconception), contributes in a major way to the shaping of these macromolecules. This step sees the emergence of **amino acids** and of catalytic centers allowing the construction of more complex metabolites. In this mineral scenario, possibly organized around iron and sulphur (common to the Earth's surface), the most important molecules are coenzymes (necessary to catalyse the chemical reactions accelerated by the enzymes that contain them), lipids (required for the construction of the cell membrane), and certain amino acids. The synthesis of lipids leads spontaneously in water (again thanks to an entropy increase) to the formation of membranes containing mineral nanoparticles, composing primitive vesicles.

These vesicles are continuously merging together and splitting, allowing both certain metabolites to be locally concentrated and to explore and share various **metabolic pathways**. The corresponding Earth's atmosphere is neutral in terms of electron transfers. This implies that the iron ion is soluble in water (this is no longer the case today because of the presence of oxygen), but raises questions about the presence of nitrogen in the molecules based on a carbon skeleton. This requirement calls for an early emergence of a mechanism for fixing atmospheric nitrogen, a very unreactive gas. This is also the time when phosphate, with polyphosphates – minerals very rich in energy but metastable in water, takes on its role of storage and chemical energy transfer, allowing the dynamic organization of metabolism. This time also witnesses the origin of nucleotides, rich in nitrogen and phosphorus. Their current metabolism is utterly different from what organic chemists do when they synthesize these molecules in the laboratory. It involves amino acids, which generate peptides, as precursors. Are there contemporary reactions synthesizing peptides outside the translation of the genetic message? Certainly, and they take place around reactions involving sulphur, precisely proposed as the origin of life by the biochemist Christian de Duve [\[3\]](#).

Non-ribosomal synthesis of peptides, peptides, polyketides

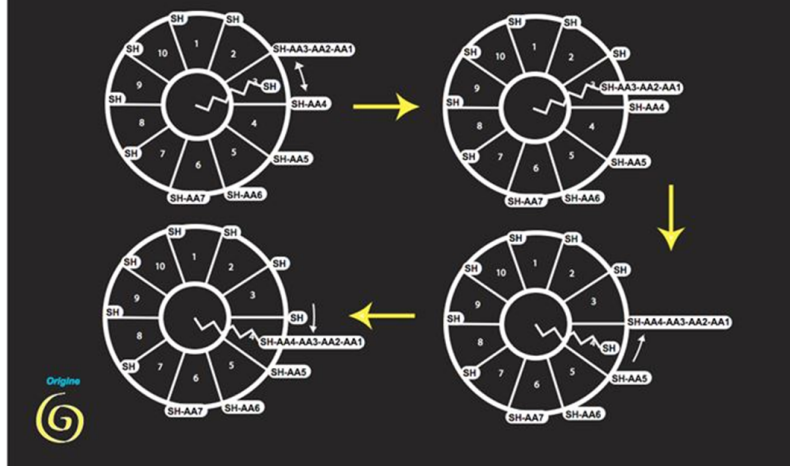


Figure 2. Synthesis of the first molecules essential to the make-up of a primitive metabolism. Fixed on a support (initially mineral) a flexible arm carrying a reactive sulphur atom displaces a substrate from one catalytic centre to another, carrying out the local polymerization of basic precursors. This is how peptides, fatty acids and more complicated molecules are still constructed today, participating in the synthesis of coenzymes and nucleotides that will lead to more and more elaborate syntheses.

These reactions produce peptides of various structures (often antibiotics). They have several archaic features: in addition to the key use of the carbon-sulphur bond, they use the two stereoisomers of amino acids (we know that life only uses the left-handed isomer in proteins). But one observation makes this hypothesis even more remarkable: the analysis of the gene sequence allowing the synthesis of fatty acids, major components of lipids, shows that enzymes belonging to the same family (and with the same coenzyme) allow their synthesis, witnessing their common origin (Figure 2). We have here a first explanation of the generation of membranes (essential for compartmentalization), as an accidental by-product of peptide synthesis.

3.2. The “RNA-metabolism” world

Based on this metabolism, a second stage is set up, that of the "RNA-metabolism" world. RNA molecules with the ability to catalyse reactions between various metabolites: these are ribozymes. Created by the polymerization of ribonucleotides, RNAs gradually replace surfaces, becoming the rigid support that allows the local variation of various substrates. The family of ancestors of transfer RNAs, adapters between the messenger RNA and proteins within the ribosome, represents the most plausible class of these substitutes. In fact, the skeleton of these molecules is today massively modified by all kinds of metabolites and there are many examples of reactions where tRNA intervenes in the reaction in an unnecessary way, unrelated to protein synthesis, witnessing a live memory of its past role. In parallel, becoming ribozymes, RNA molecules have gradually discovered how, like protein enzymes, they could catalyse reactions between various metabolites. In this context, the formation of peptides, initially random, appears to be a key reaction retained by the RNA molecule that would become the ribosome. The ancestor of the ribosomal ribozyme initially used tRNA ancestors as amino acid holding devices in the process forming a peptide bond. The specificity of the amino acid sequence in peptides subsequently materialized when an RNA matrix imposed a strict succession order to the amino acid-loaded tRNAs. Finally, the three-dimensional folding of the RNA on itself, necessary for the generation of ribozymes, led to the discovery of the law of sequence complementarity (formation of a double helix in which nucleotides are complementary). Then, this law of complementarity, which associates a particular tRNA with a sequence of the RNA template, gradually crystallized into the form of a rigid correspondence between triplets of nucleotides and amino acids, giving rise to the amino acid-nucleotide cipher, generating the rule that forms the genetic code.

3.3. Invention of replication: “RNA-genome” world

Vesicles containing coding RNA molecules repeatedly split and fused, reproducing and propagating increasingly efficient metabolic pathways in parallel with the associated peptide-coding assembly system. This revolution changed the course of things: an "RNA-genome" world, based on RNA as a (replicable) information template got separated irreversibly from the RNA-metabolism world. The law of complementarity allowed the formation of a copy of the RNA template used for peptide generation, in the form of an RNA double helix. The emergence of peptides promoting its replication opened up the world of RNA genomes, where transcription of RNA and replication a process for obtaining two molecules identical to the original molecule. A transcription process produced templates for decoding the message by the protoribosome, while replication increased the number of copies of this RNA ancestor of genes. This step led to the formation of vesicles containing sets of various sequences of RNA double helices, as found in the genome of some RNA viruses today. However, these protocells depended on the ongoing synthesis of

ribonucleotides and polyribonucleotides, which are chemically very unstable molecules. This continuous process required the association within the same compartment (a primitive cell) of both the reproduction of the RNA-metabolism world and the specific replication of the RNA-genome world. The RNA-genome world had therefore to find a way to stabilize the synthesis of its precursors so as not to be doomed to disappear. The discovery of deoxyribose, much more stable than ribose, resolved this quandary. Thus, DNA, a molecule very stable over time, appeared only at the very end of the process that resulted in memorizing the recipes that drive the production of metabolism. This novel molecule kept memory apart from general functioning of the cell. Subsequently, DNA copies of the coding RNAs fused together, forming the first chromosomes.

4. The first cells



Figure 3. The first cells were phagocytes. These cells, forming a cytoplasm (derived from the RNA-metabolism) and a nucleus (issuing from the RNA-genome, later substituted by DNA), engulf other cells of the same type. This allows them to capture elements essential to their metabolism but also to recover DNA fragments encoding functions invented elsewhere. "Horizontal gene transfer" is therefore the rule at this stage of evolution.

This final stage witnesses the moment when the first cells were born. To understand their emergence, it is necessary to explore how vesicles surrounded by a lipid bilayer interact. They can split and merge, but a constant property is their ability to interpenetrate, as observed today in the process of phagocytosis. Process allowing a cell to encompass and then digest a foreign substance or organism (e. g. bacteria). This process considerably enriches the evolution of metabolic systems since it allows the association in a single cell of compartments with different but complementary destinies. The ancestor of the first cells was a phagocyte, associating an RNA-metabolism -ancestor of the cytoplasm Internal cell environment. It consists of a water- and protein-rich phase (cytosol) and contains cellular organelles (mitochondria, etc.) with translation of the genetic message and intermediate metabolism, and an RNA-genome -ancestor of the core genome in particular after the emergence of DNA and the separation of the cell's metabolism into replication and transcription into RNA.

At this point of evolution, we are therefore faced with a set of phagocytes, the **protokaryotes**, which evolve by reproducing and systematically ingesting what surrounds them. They are fairly large cannibal predators (like protists today), producing metabolic innovations that they consume and propagate.

This situation is unstable. Indeed, it brings out a specific function, the one that allows resistance to phagocytosis. If an organism with a flexible metabolism finds a structure that can carry out this resistance, it will escape the cannibalism of protokaryotic organisms and start a new evolutionary lineage.

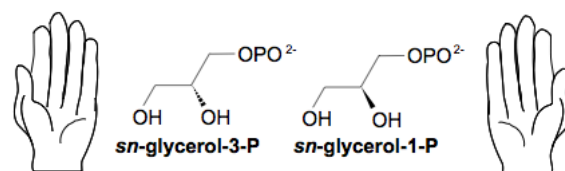


Figure 4. Skeleton of membrane lipids. Membrane lipids combine long-chain fatty acids with molecules such as glycerol phosphate. Eukaryotes and Bacteria use a different stereoisomer of this backbone than that used by Archaea. The result is that while the membranes of the latter can in a limited way merge with those of the former, the protein moiety that belong to them (transporters, shields etc.) cannot.

Two solutions to this barrier are possible: surround the cell with an envelope that is very difficult to ingest, or make phagocytosis impossible for physico-chemical reasons. Bacteria are the descendants of the cells that have found the first solution, forming

small cells and surrounded by a resistant envelope. Archaea are those which have discovered how to surround themselves with a membrane that is impossible to ingest as a functional structure, using lipids with a three-dimensional structure that mirrors the lipids of their predators. They further escaped by colonizing extreme environments (Figure 4).

To sum up, the emergence of stable **deoxyribonucleotides** has allowed the grouping of genes within chromosomes, while phagocytosis has led to an escape process based on a metabolic alternative to the structure of membrane lipids and the conquest of extreme environments, with Archaea. Single-celled prokaryotic microorganisms living in particular in extreme environments (anaerobic, highly saline, very hot...). Phylogenetic research by Carl Woese and George E. Fox (1977) differentiated between archaea and other prokaryotic organisms (bacteria). Currently, living organisms are considered to consist of three groups: archaea, bacteria and eukaryotes, and on the emergence of a robust and resistant envelope, with **Bacteria**. Then, as Bacteria and Archaea miniaturized the cell, some Bacteria regained an interest in sharing metabolism, simplifying their envelope so that some of them became symbionts of protokaryotes (in a way reminiscent of the way some bacteria form nodules in plant roots). The pursuit of a reductive evolution transformed this symbiosis to the point of reducing the bacterial genome to a small set of genes, within **mitochondria**, organelles specific to eukaryotes. Unicellular or multicellular organisms whose cells possess a nucleus and organelles (endoplasmic reticulum, Golgi apparatus, various plasters, mitochondria, etc.) delimited by membranes. Eukaryotes are, along with bacteria and archaea, one of the three groups of living organisms. (see [Symbiosis and evolution](#)).

Note that this scenario could possibly be validated experimentally. This would be the case if somewhere there was still a direct descendant of the protokaryotes. This hypothetical organism would have the particularity of having a cytoplasm and a nucleus, but no mitochondria (nor traces of organelles of bacterial origin). Since the main function of mitochondria today is the formation of iron-sulfur clusters essential for the activity of many enzymes, these organisms must be sought in environments where access to these structures is easy. If such an organization existed, it would be a considerable asset to transform the present scenario from an educated guess into a scientific reality.

5. Towards a possible scenario of life's origin

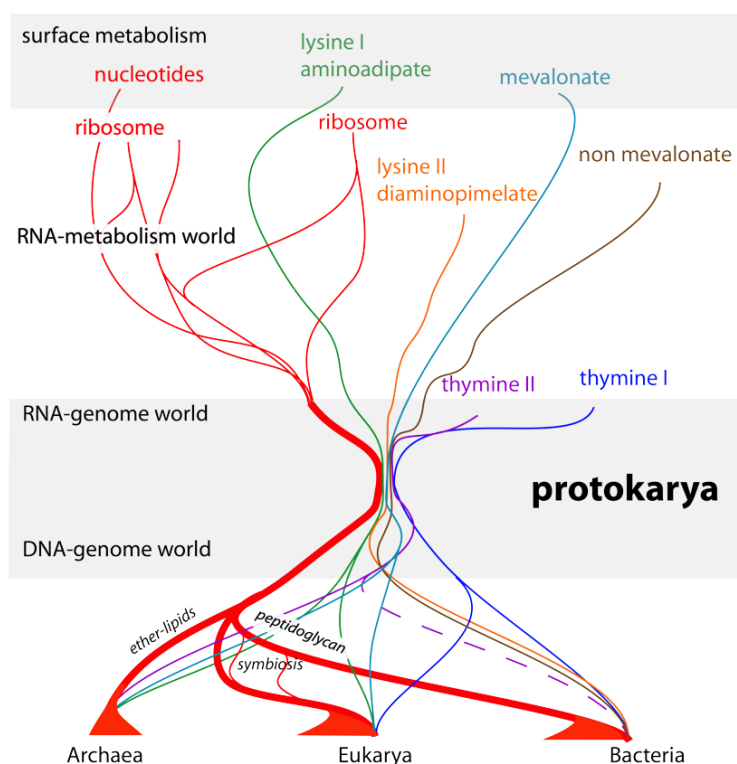


Figure 5. A scenario of the origins of the first cells. Following a surface metabolism step, which ended up in the emergence of the main metabolites (in particular, besides amino acids, nucleotides, coenzymes and lipids), two RNA worlds (the RNA-metabolism world and the RNA-genome world) joined forces to create the ribosome ancestors, coupling these two worlds. Several synthesis pathways led to the protein amino acid lysine, the DNA base thymine, and isoprene derivatives that produced important coenzymes and lipids. After the discovery of DNA, genes were grouped together into chromosomes, and the first cells, protokaryotes, evolved from their phagocytic behaviour. The emergence of a particular pathway for lysine synthesis (using diaminopimelate as precursor, instead of amino adipate) allowed some of these cells to form a phagocytosis-resistant envelope. Similarly, the synthesis of phospholipids from a different glycerolphosphate isomer than the one previously present allowed the ancestors of the Archaea to escape phagocytosis. Current eukaryotes appeared after Bacteria were ingested but not killed to live in symbiosis with their protokaryotic host.

In summary, the proposed **scenario of origins** (Figure 5) assumes that individuals of the same species are all different from one another. However, it also assumes that it is a common **program** that decides on their construction. This program is transmitted from generation to generation, unmodified (it replicates), while individual cells only reproduce (they are similar, not identical, to one another). This distinction between **reproduction** and **replication** is essential to understand what life is, a dialogue between the inevitable production of **variants**, and the maintenance over generations of a program that is as invariant as possible.

At the origin of the first cells, a **flow of chemical reactions**, a primitive **metabolism**, reproduced, generating all kinds of hopeful accidents but far too many errors to remain sustainable. Some, later, gave rise to the exact replication of a structure associated to its functioning via a coding process. **Metabolism** began in **water** on the **surface of minerals**. Surfaces, unlike what would happen in a primordial soup, not only helped retaining only a tiny fraction of all chemical inventions centred on the carbon atom, but also promoted the **polymerization of** some of these **building blocks**, thanks to an entropy-driven process involving water.

Some of these building blocks gave rise to macromolecules, **RNAs**, that substituted to the surface of minerals. An **RNA-metabolism** world developed in this way. It then led to the discovery of the **replication of** these RNAs, forming the **RNA-genomes** world. The latter used **RNA** as an **information template** (primitive genes) rather than as a direct substrate for metabolism.

Thanks to the invention of molecules to **separate an internal environment from the outside**, **membrane lipids**, the **first cells** (protokaryotes) brought together two compartments, one from the RNA-metabolism world (**cytoplasm**) and a second one from the RNA-genome world (the **nucleus**). These cells were **phagocytes** that devoured everything they met, and thus spread metabolic **innovations** very quickly. Subsequently, a variant of RNA, the notorious DNA, allowed the grouping of genes within **chromosomes**.

In a final twist, **phagocytosis** opened up the possibility of an escape based on the alteration of the cell's envelope. Two escape ways were possible. Membrane lipids are key to allow phagocytosis, but they are asymmetric. Replacing them with their mirror image reduces or eliminates phagocytosis. This change in membrane symmetry triggered the origin of the **Archaea**. Covering the lipid membrane with a very resistant structure was another solution, and this decided of the origin of the **Bacteria**. However, phagocytes still found a way to ingest some of these without killing them altogether, and this triggered the origin of **eukaryotes**, which include the plants and animals as we know them.

Notes and references

Cover image. HeLa Cells [Source: © National Institutes of Health (NIH), via Wikimedia Commons]

[1] Ockham's razor: principle of philosophical reasoning, also called the principle of simplicity or parsimony.

[2] Freeman Dyson (1986) *Origins of Life*, Cambridge University Press. Second edition 1999 (ISBN 0521626684)

[3] Christian de Duve (1917-2013), Belgian biochemist who received the Nobel Prize in 1974 for the discovery of previously unknown organelles in the cell, lysosomes. These have important functions in decomposing different types of materials, such as bacteria and parts of cells that have worn out.

L'Encyclopédie de l'environnement est publiée par l'Université Grenoble Alpes - www.univ-grenoble-alpes.fr

