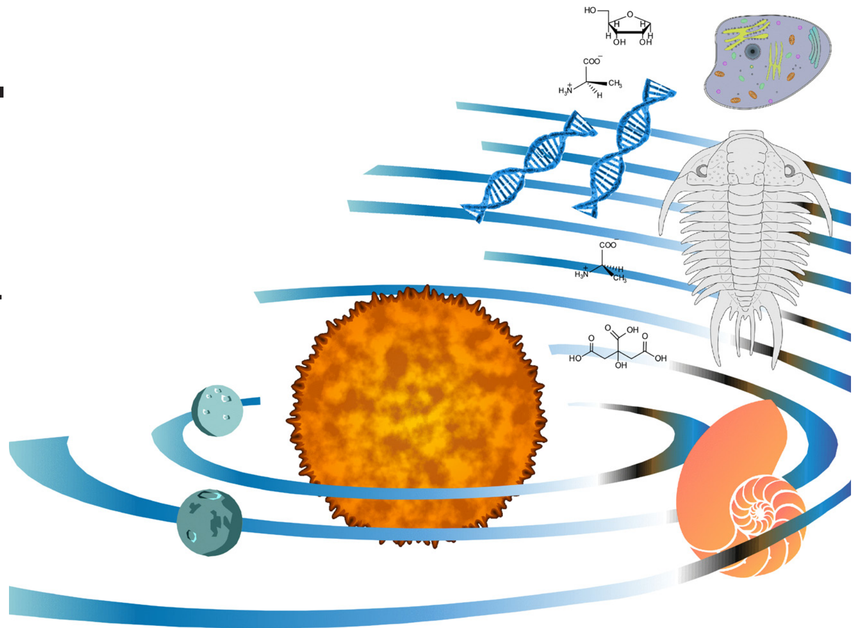
The first self-replicating molecule and the origin of life



Roberto Laos and Steven Benner*

Accede a la versión en español aquí Descargar

The origin of life on Earth is one of the most challenging questions in science. In the last sixty years, considerable progress has been made in understanding how simple molecules relevant to life can be generated spontaneously or are known to arrive to Earth from space. Additionally, analysis of the evolution history of nucleic acids, which are the repository of genetic information, points to a now extinct, universal common ancestor for all life on Earth. The studies of the origin of life offer many clues towards a common origin, perhaps not just on Earth but somewhere else in the solar system. However, due to the length of time that the Earth has harbored life, the oldest clues of the first organisms are mostly gone. It is unlikely to find exactly what this first organism was like. Nevertheless, in the last few years, synthetic biology has made remarkable progress at modifying biomolecules, particularly nucleic acids. It is possible that soon we will be able to construct and understand a minimalistic system in which molecules can copy themselves in a protocell. The study of such systems could shed light onto the origin of the first organisms.

Keywords: Origin of life, synthetic biology, prebiotic chemistry, RNA world, AEGIS

El origen de la vida en la Tierra es una de las preguntas más difíciles presentadas a la ciencia. En los últimos sesenta años, ha habido un progreso considerable en entender cómo moléculas relativamente sencillas, que son relevantes para la vida, pueden ser generadas espontáneamente o pueden llegar a la Tierra desde el espacio. Además, los análisis de la evolución de la historia de ácidos nucleicos, los cuales almacenan la información genética, apuntan a un ancestro común universal ya extinto. Los estudios del origen de la vida ofrecen muchas pistas que apuntan hacia un origen común, quizás no solo en la Tierra sino también en algún otro punto del sistema solar. Debido al largo tiempo transcurrido desde que la Tierra empezó a albergar vida, las pistas más antiguas de los primeros organismos se han perdido. Es muy poco probable encontrar exactamente cómo fue este primer organismo. Sin embargo, en los últimos años la biología sintética ha logrado progresar mucho en la modificación de biomoléculas, en particular, los ácidos nucleicos. Es posible que pronto podamos construir y comprender un sistema minimalista en el cual las moléculas puedan copiarse a sí mismas dentro de una célula rudimentaria. El estudio de un sistema así podría permitirnos develar el origen de los primeros organismos.

Palabras clave: Origen de la vida, biología sintética, química pre biótica, mundo ARN, AEGIS

* *Roberto Laos (B.Sc. in Chemistry from PUCP) and Steven Benner (PhD in Chemistry from Harvard University, U.S.A.), have published several papers in prestigious peer reviewed international journals on molecular biology. Moreover, Proffessor Benner is the founder of the Foundation for Applied Molecular Evolution (FFAME), a prestigious north american institution located in Gainesville, Florida, where the authors belong. (e-mail (Roberto Laos): rlaos@ffame.org)*

In the year 1953, two publications set the foundations of what would be later the studies of the origin of life and would help us better understand the nature of living beings. With time, these articles would complement each other. One of them reported the structure of DNA by Watson and Crick,[1] who closed their article in *Nature* by pointing out that the double helix structure of DNA could be fundamental for the copying and transferring of genetic information. The other, published in *Science*, was by Stanley Miller.[2] Miller discharged electricity through a mixture of methane, ammonia, and hydrogen, simulating what lightning would do to these gases on early Earth. He recovered several amino acids, essential components of living organisms that we know.

Both of these discoveries opened a series of new disciplines. The studies of Miller mark the beginning of the studies of the origins of life as an experimental discipline. Nowadays, more than sixty years after these publications, considerable progress both in chemistry and biology has been accomplished as well as in geology, space exploration, and molecular biology. Even new disciplines like astrobiology and synthetic biology have appeared. This large body of knowledge has helped us to understand that the living beings on Earth have a lot in common, not just at the molecular level but also in our shared common past.

Currently, the Earth is considered to have harbored life for at least the last 3500 million years (3.5 Ga)*. This time scale is difficult to grasp for a person who is lucky to live 100 years. It is convenient to put this in perspective using the "cosmic calendar" popularized by the late Carl Sagan.[3] In this calendar, the history of the world is scaled down to one year. The reference points are the big bang, which occurred on the first second of January 1st, and our present time, which occurred on the last instants of December 31st. According to this scale, the conquest of the Inca Empire occurred a second ago. Our lives are a blink of an eye on this gigantic scale. Following this scale, the formation of the solar system happened in September and life on Earth appeared in October. The relative speed with which life apparently arose suggests that life can arise fairly easily. We will return to this point later.

The study of the origin of life might sound like an idealistic crusade that would only be used to satisfy our natural curiosity; however, there are several practical aspects in studying alternative nucleic acids, ancient proteins and to explore the space around us.

On broad strokes, the origin of life can be described the following way: After the big bang, stars began making heavier chemical elements by fusion of hydrogen and helium atoms. Some of the stars became supernovae and formed even heavier elements, those heavier than iron. Our solar system was formed 10 Ga years later from material that had, on average, passed through three or four supernovae. The accretion of this material, to give us the Sun, the Earth and the other planets in our solar system was triggered when a nearby star, about 25 light years away, itself went supernova; the resulting shock wave disrupted the gas cloud and caused it to collapse.

Soon after it was formed, the Earth was hit by a Mars-sized object. This colossal crash formed the Moon, our natural satellite. If there were life before this impact, it is impossible that it would have survived and also impossible to find any trace of it.

After the formation of the Moon, Earth cooled quickly for about 600 million years. Then, about 2.8 Ga years ago, both, the Earth and the Moon suffered what is known as the late heavy bombardment; during this period an unusually large number of meteorites hit the planet. The marks of this bombardment are the numerous craters that can be seen easily in the Moon. It is not clear whether life, had it emerged rapidly on Earth, would have been extinguished by the late heavy bombardment. What is clear is that the record of life on Earth, nearly absent before the bombardment, becomes recognizable soon after the bombardment. At 3.5 Ga ago, Earth almost certainly had life, and life having approximately the same molecular core as modern life on Earth.

How did this happen? In a model suggested by Charles Darwin, organic molecules accumulated in a "warm little pond,"** and then somehow linked themselves to form molecules able to direct their own replication, where the replication was imperfect, and the imperfections are themselves replicable. These are the elements of Darwinian evolution, which is assumed to be the only way the properties that we value in life might emerge. These self-replicating molecules were eventually encapsulated, and the first primitive cells arose. Finally, the first modern organisms appeared, with the last 3.5 Ga allowing Darwinian evolution to give the many different species that we have on Earth today.

A definition of life

The first step in our quest to understand the origin of life, or to find other life forms, is to define what we are looking

* 3500 million years equal to 3.5 billion years or 3.5 Ga. Different units of time are used in the literature such as My and May; in this review we use the prefix giga. 1Ga = 100 000 000 years.

** This is exactly as Darwin describes it in a personal letter to Joseph Hooker. The letter is dated February 1st, 1871.

1. Watson, J. D., Crick, F. H.C., Nature, **1953**, *171 (4356)*, 737-738.
2. Miller, S. L., Science **1953**, *117 (3046)*, 528-529.
3. Sagan, C., Natural History **1975**, *84 (10)*, 70-73.

Figure 1. Fossil of Leptomeryx. This mammalian ruminant herbivore, is an ancestor of the deer, giraffe, antelope, and ox and lived in North America 40 million years ago during the Oligocene. From the S. A. Benner collection; photographed by himself.

for. One approach to define life lists properties of the life that we know. For example, Daniel Koshland, editor of the prestigious American journal Science, found seven properties of the life we know: "program," meaning genetic structures; "improvisation," meaning the ability to change structure as the environment changes; "compartmentalization," meaning that life is separate from non-life; "energy," meaning that life consumes resources that are not at thermodynamic equilibrium; "regeneration," meaning that life has children; "adaptability," meaning the ability of the system to respond to changes, and "seclusion," meaning that life has physically separated chemical reactions within it.[4]

Such "make a list" approaches to the definition of life easily find counterexamples. For instance, fire certainly consumes fuel ("energy"), regenerates itself to have "children," and changes as the environment changes. However, we know that fire is not alive. Fire is missing not only the program, but the ability of its descendants to be different from itself, and for those differences to themselves be passed on to the next generation.

... life would be defined as *"a self-sustaining chemical system capable of directing its own reproduction, where the replicates are imperfect, and where the imperfections are themselves replicable".*

An alternative definition of life focuses on the mechanism by which life improves itself: by making heritable variation that is then selected naturally. This is the basis for a 1994 definition of life as "a self-sustaining chemical system capable of Darwinian evolution." [5] Here, much of the meaning is captured in the phrase "Darwinian evolution." This includes replication, but also where the replicates (your children) are not exact (because of mutation). It also allows the mutations that distinguish your children from you to be passed to your grandchildren if they confer "fitness." This second step is essential for Darwinian evolution; it is what allows the genetic program to adapt, evolve, and get better over time.

But even here, the definition has problems. First, Darwinism requires that the mutations be random. Especially, the mutations cannot be "prospective" with respect to fitness. Darwinian organisms cannot know what the future holds, and therefore, cannot create mutations in anticipation of future needs. Darwinian mechanisms specifically exclude the possibility that an organism might deliberately create specific mutations that will make its descendents more fit.

Human technology, of course, is undermining this concept. We can imagine a future in which we will still direct our own reproduction, but deliberately make our replicates imperfect by adding mutations that we know will help our children survive. Perhaps we will do this on Mars, where the environment might demand dramatically different genes. Such a process is excluded by Darwinism. Thus, if we insist that life is simply "a self-sustaining chemical system capable of Darwinian evolution," the human lineage that invents this technology will magically cease to be "life".

Accordingly, a better definition replaces the concept of "Darwinian evolution" with its particulars. Here, life is "a self-sustaining chemical system capable of directing its own reproduction, where the replicates are imperfect, and where the imperfections are themselves replicable".

What do we know? What can we know? What we will never know?

The history of Earth can be read in the geological record and the fossil record. Both offer abundant information, and both have its limitations. All the rocks on Earth go through a cycle of erosion and deposition. As a consequence of this cycle the oldest rocks are the scarcest. Since the Moon has no volcanic activity and once was part of the early

4. Koshland, D. E., *Science* **2002**, *295 (5563)*, 2215-2216.
5. Deamer, D. W.; Fleischaker, G. R., "*Origins of life: The central concepts*". Jones & Bartlett Publishers, 1994.
6. Mojzsis, S. J.; Arrhenius, G.; McKeegan, K. D.; Harrison, T. M.; Nutman, A. P.; Friend, C. R. L., *Nature* **1996**, *384 (6604)*, 55-59.

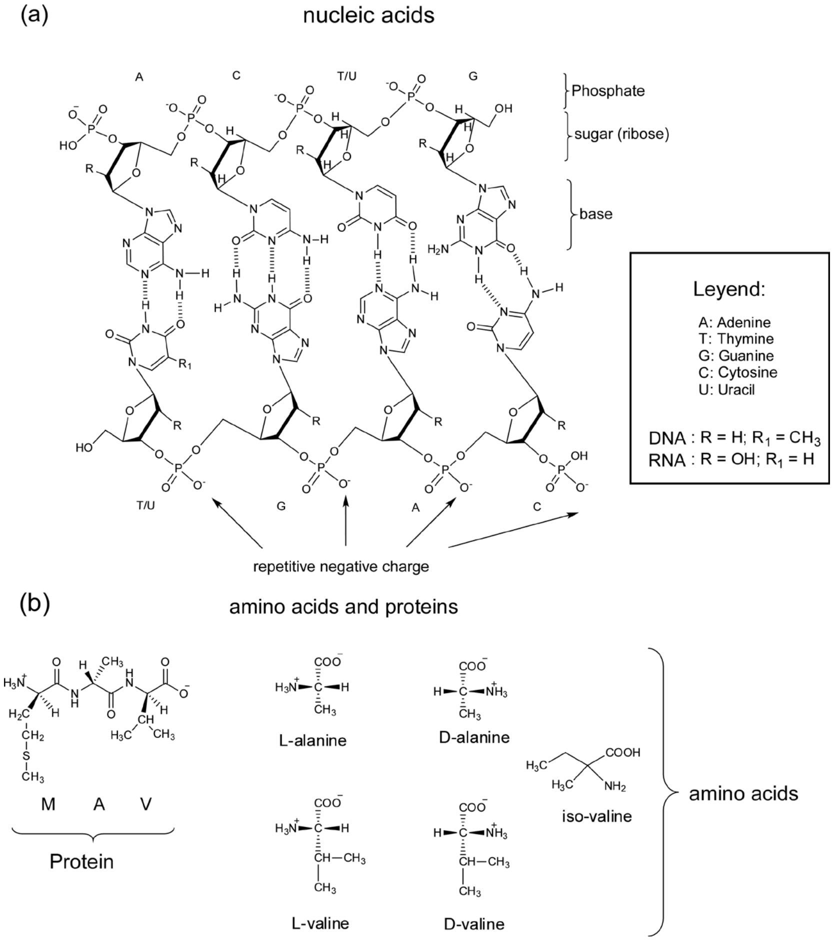


Figure 2. Biomolecules on Earth. Nucleic acids, DNA and RNA are formed by a phosphate, a sugar and a base. The bases pair with each other by hydrogen bonding (weak interactions between O/N and H). The rules of pairing are: A pairs with T (or U on RNA) and G pairs with C. This is observed through all biology found on Earth. Chains of amino acids form proteins like the one depicted: M-A-V. Amino acids are chiral molecules. The same atoms can be arranged differently in space. Some of the amino acids found in the Murchinson meteorite (Figure 3) are shown here: D-alanine, L-alanine, D-valine, L-valine, isovaline. However only L-alanine and L-valine are found to be parts of proteins on living organisms.

Earth, this is the place where we can find rocks that are as old as the Earth.

The fossil record is the collection of fossils, living beings that were buried, to be later dug up. Often, their bodies have been mineralized, leaving only their shapes (Figure 1). The oldest evidence of life on Earth comes from some fossils found in the Isua Greenstone Belt,[6] and these are at least 3.8 Ga old. Microbial life may have existed as far back as the terrestrial rock record extends. This may lead us to think that life can arise relatively easily. One of the deficiencies of the fossil record is that it does not provide a representative sample since some species fossilize better than others. Because of this, we have more fossils of hard-shelled animals. Moreover, it is not a representative sample geographically, either, since the conditions for fossilization are not found evenly through the planet. For instance, the Peruvian jungle is not a good place for finding a fossil since a dead animal is more likely to be devoured by another animal or insect or to be washed away by rain before it is buried and mineralized to form a fossil. Nevertheless, the fossil record together with the geological record document massive extinctions that happened on Earth. On these events, the large majority of living beings suddenly disappeared. This is seen by the sudden absence of certain fossils. Perhaps the most commented massive extinction is the one that involved the extinction of dinosaurs, which is related to the fall of a large meteorite on the Yucatan peninsula.

Biochemistry on Earth is universal

All organisms known to man work the following way: genetic information is stored in DNA, a polymer formed by four nucleotides: A, T, G, C. This DNA is copied by an enzyme called RNA polymerase,* which produces another polymer called RNA, formed by the nucleotides: A, U, G, C. The RNA is taken to the ribosomes where RNA is read and the information is translated to proteins, polymers made out of 20 different amino acids, which can be represented by the letters: M, A, D, E, G, F, L, S, Y, C, W, P, H, Q, R, I, T, N, K, V. For instance, a sequence of DNA like: ATG GCC GTA will translate to a protein with the sequence: M A V.

This information transfer from DNA to RNA to protein is known as the "central dogma" in biology because it is observed by all known organisms. This transfer of information occurs with great speed and accuracy. Even the simplest unicellular organism is complex enough to be able to produce biomolecules with speed and accuracy better than any man-made laboratory. What is more remarkable is that all living beings use the same set of biomolecules from a large collection of possible alternatives. Furthermore, these biomolecules have a property know as chirality. Two molecules are called chiral when they differ only by the spatial arrangement of their atoms. Chiral molecules are non-superimposable mirror images of each other, like the right and left hand. All the biomolecules used by living systems that we know (Figure 2) use a set of molecules, and they have cer-

* DNA and RNA are the acronyms for deoxyribonucleic acid and ribonucleic acid, respectively.

\# Ribosomes are composed of RNA and amino acid chains (proteins).

Figura 3. The Murchinson meteorite is a carbonaceous chondrite that fell on Australia in 1969. This kind of meteorite has a dark color that comes from the organic compounds it contains. It is one of the most studied meteorites. From the S. A. Benner collection; photographed by himself.

tain chirality. This surprising uniformity of life on Earth points to the fact that all living beings have had a common ancestor.
.

Prebiotic chemistry

We call prebiotic chemistry the reactions that happen in the absence of living organisms, especially those that might have occurred on early Earth. Others of these reactions may have occurred elsewhere in the cosmos, with their products being brought to Earth by meteorites. These meteorites are fragments produced by the collisions that happen in the asteroid belt. When a meteorite enters the Earth's atmosphere, its surface melts, but its interior remains unaltered. These fragments contain frozen samples of the formation of the solar system. A particular kind of meteorite, called carbonaceous chondrite, is of special interest (Figure 3). These chondrites contain organic molecules, such as amino acids. Some of these are identical to those found today in terran proteins; others are not.

Stanley Miller is considered a pioneer of modern prebiotic chemistry. The experiments of Miller demonstrate that many of the molecules found in modern biology, particularly amino acids, can be produced without the intervention of living organisms. Assuming a strongly reducing atmosphere for the early Earth,[7] he exposed a mixture of CH_4, NH_3 and H_2 (a reducing gas mixture) to an electric discharge, recovering from the tarry products (after acid digestion) amino acids.

In the years following these experiments, the synthesis of biologically interesting molecules from products that could be obtained from a reducing gas mixture became the central aim of prebiotic chemistry. Oró and Kimble were able to synthesize adenine from hydrogen cyanide and ammonia.[8] Later, Sanchez, Ferris, and Orgel[9] showed that cyanoacetylene is a major product of the action of an electric discharge on a mixture of methane and nitrogen, and that cyanoacetylene is a plausible source of the pyrimidine bases uracil and cytosine.

This new information, together with previous studies that showed that sugars are formed readily from formaldehyde by the Butlerov reaction,[10] suggested that the first stage in the appearance of life on Earth involved the formation of a "prebiotic soup" of biomonomers. In the early Eighties, the demonstration that ribosomal peptide synthesis is a ribozyme-catalyzed reaction[11] reinforced the idea that there was once an RNA World. One of the central problems for the origin of life studies is then to understand how a protein-free RNA World became established on the primitive Earth. This idea of the RNA world will be discussed further. Currently, there is no complete reaction that can produce nucleic acids (DNA, RNA) which are the most complicated molecules, formed by three distinct subunits: ribose (a sugar), phosphate, and a nitrogenated base (Figure 2).

One common ancestor for all

Bioinformatics help us to analyze sequences of the three fundamental biomolecules. Linus Pauling and Emile Zuckerkandl[12] were the first to propose that it was possible to infer the sequences of proteins of extinct organisms by comparing the sequences of their descendants. Using sequences of the ribosomal RNA, it is possible to construct a phylogenetic tree (similar to a family tree) of all known organisms. In this tree it can be seen how all living organisms are placed into three main groups: Archaea, Bacteria, and Eukarya. Each of these main branches of the tree comes from a common ancestor that is already extinct, the last universal common ancestor (LUCA). The understanding of this com-

7. Oparin, A. I., "*The origin of Life*". Macmillan: New York, 1938.
8. Oro, J., *Biochem. Biophys. Res. Commun.* **1960**, *2 (6)*, 407-412. (💻)
9. Ferris, J. P.; Sanchez, R. A.; Orgel, L. E., *J. Mol. Biol.* **1968**, *33 (3)*, 693-704. (💻)
10. (a) Butlerow, A., *Compt. Rend. Acad. Sci.* **1861**, *53*, 145-147; (b) Butlerow, A., *Liebig's Ann. Chem.* **1861**, *120*, 295.
11. Cech, T. R.; Zaug, A. J.; Grabowski, P. J.; Brehm, S. L., *Cell Nucleus* **1982**, *10*, 171-204. (💻)
12. Pauling, L.; Zuckerkandl, E., *Acta Chem. Scand.* **1963**, 17, s9-s16. (💻)

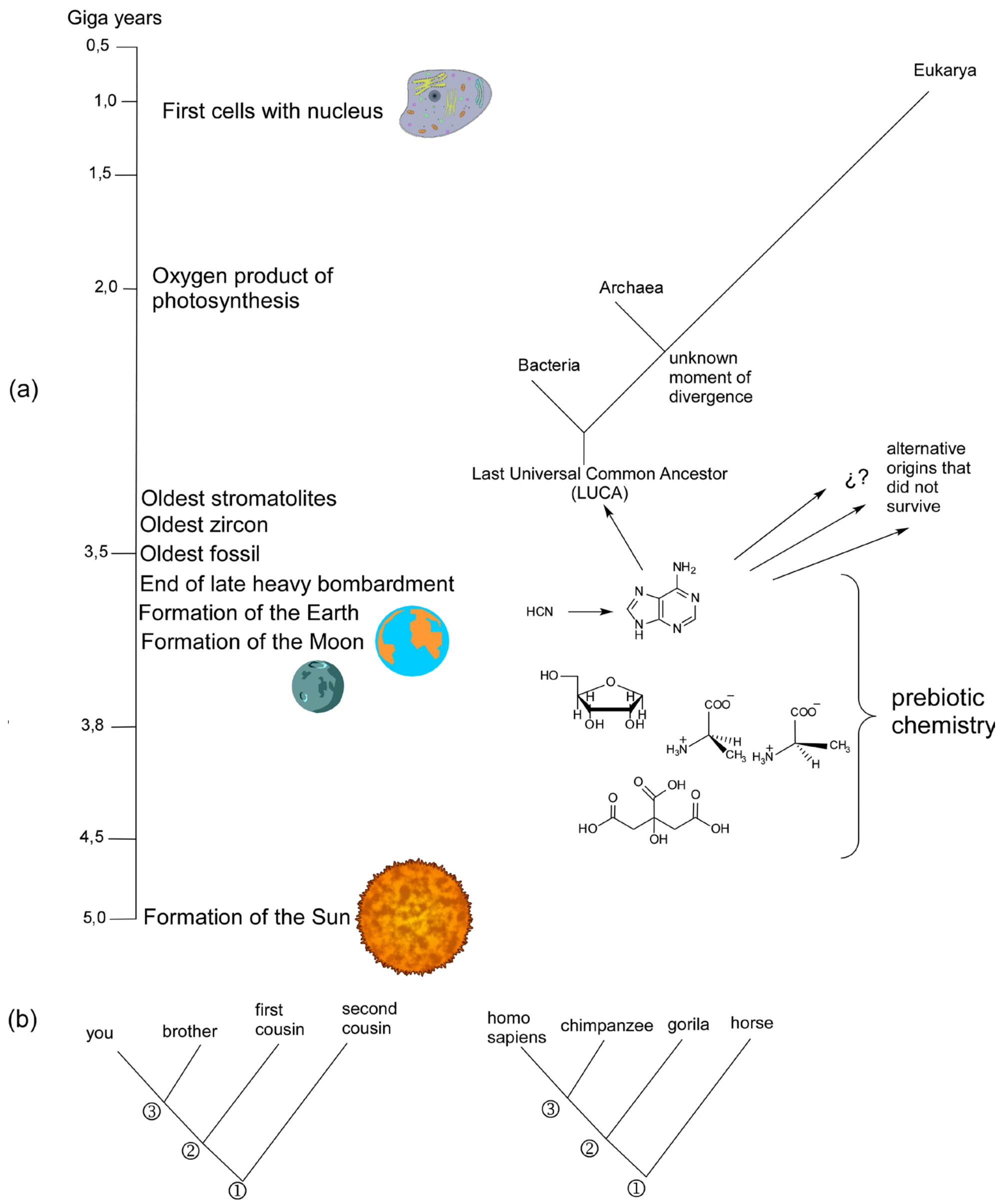


Figura 4. (a): Time-line with some events relevant to the history of the Earth. Time of appearance of the Last Universal Common Ancestor (LUCA) and the eukaryotes are known. However, the time when Archaea diverged from Bacteria is not clear to this date. The study of the origin of life is focused in unraveling how small molecules assembled and started to replicate.
(b): Phylogenetic trees compared to a family tree. Looking on your father's side, you and your brother have your father as your earliest common ancestor (3). You and your first cousin share a grandfather as your first common ancestor (2). The last common ancestor you share with a second cousin is your great-grandfather (1).

mon ancestor would give us a good idea of what differentiated this organism from unanimated matter. Figure 4 shows a simplified version of this tree parallel to a time line that marks key events on the history of Earth. The appearance of the common ancestor happened at least 3.5 Ga ago. The apparition of eukaryotes (their cells have a nucleus) happened approximately one Ga ago. Among the eukaryotes are all animals and plants.

Before the last common ancestor, the RNA world

What is earlier than the LUCA? What can be simpler than the simplest organism?

First, LUCA was not "simple" in any chemical sense of the term. It was able to make three different biopolymers and have those interact with each other following the central dogma of biology. Any such organism is already too evolved and complex to have appeared spontaneously from prebiotic reactions. Is it possible that one of these molecules appeared before the others?

In our biochemical system, each molecule has a defined role. Information is stored in DNA and transferred to RNA. The specialized proteins, called enzymes, are in charge of the catalysis of chemical reactions in living beings. Then we face a paradox similar to the chicken or egg. Which molecule appeared first? DNA or proteins?

Some think that proteins came first since Miller demonstrated that it is possible to produce amino acids, the building blocks of proteins in prebiotic conditions. Others think that DNA came first since genetic information is stored on DNA. However, no full prebiotic route is known for the formation of DNA. The discovery that RNA can act both as catalyst and as repository of genetic information gave huge support to the idea that there was once an RNA world where there were no proteins and no DNA. Even though molecules do not leave fossils, we can see remnants of this RNA world in modern metabolism. For instance, molecules like ATP, acetyl co-enzyme A, FAD, thiamine, and NAD^+ contain

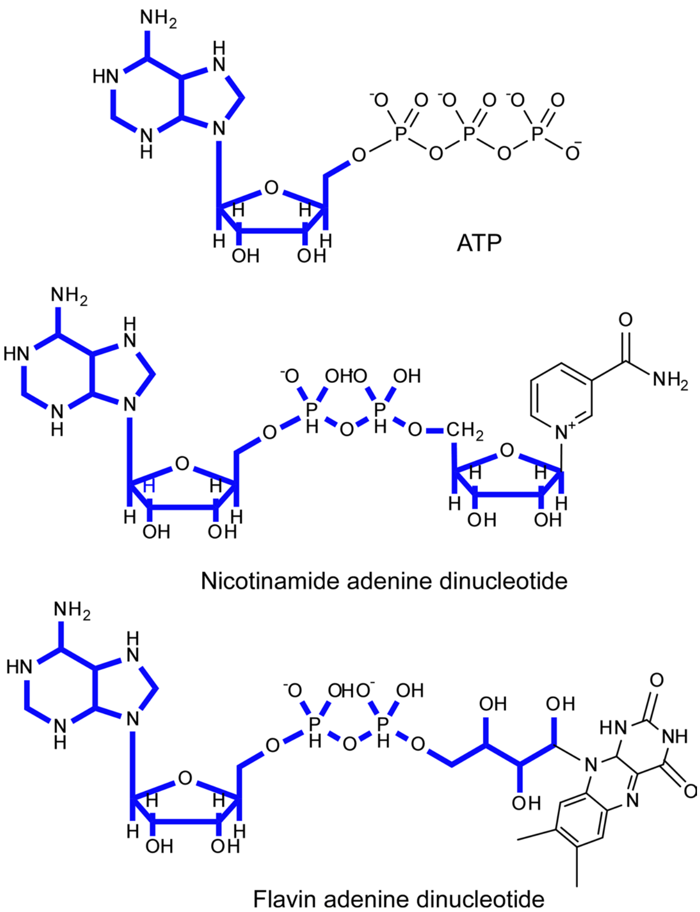


Figure 5. Molecular fossils. Cofactors involved in metabolism contain RNA portions (shown in blue) attached to reactive portions (shown in black). The RNA portions do not participate in metabolism and are believed to be vestiges of a time on Earth when life used RNA as the only encoded biopolymer.

fragments of RNA, which are non-essential for the reactivity of these molecules. The simplest explanation (which tends to be the correct explanation) is that these molecules were part of metabolic processes in the RNA World, and our modern world is like a palimpsest where the remnants of the RNA world can still be seen in some molecules[13] (Figure 5).

What will we never know?

We will never know, without a doubt, which was the first self-replicating molecule. Neither will we know which was the first living being because it is possible that more than one origin of life happened, but if this happened before the collision that formed the Moon, we have no chance to find a trace of this life.

Four approaches to understand life

We can now review four approaches that can lead us to understand the origin of life. The first goes backwards in time, using biotechnology, from modern life towards the resurrection of ancient genes and proteins to the study of them in the laboratory. The second works forward in time, starting with simple molecules that can be formed in the absence of living beings, to ask questions on how these molecules gave origin to the first living systems. The third considers unusual environments in our solar system, particularly those that deviate from the environments that harbor life on Earth. The last involves synthesis, in which Darwinian systems able to replicate themselves are built artificially in order to test the biological properties that they can produce. Together, these four approaches are constraining the "black box" that captures the phenomenon of life.

(1) Backwards in time: bioinformatics

Nowadays, a large amount of sequences of both proteins and nucleic acids are available and can be used to infer the sequences of proteins of extinct organisms and understand better the ancient life that once inhabited the Earth. Studying the proteins of ancient organisms can help us to understand better the common details that these extinct organisms had in common. These details are buried under 3.5 Ga of evolution. For instance, enzymes from of ancient bacteria have been found to have an optimal temperature higher than their modern counterparts.[14]

(2) Forward in time: prebiotic chemistry

Previous attempts of producing sugars in prebiotic conditions lead to intractable mixtures. The Butlerov reaction, also known as the formose reaction, produces an intractable mixture of sugars in which ribose is usually only a minor product.[15] Although a mechanism consistent with experimental evidence can be written for the formose reaction[16], until recently it had been not possible to channel the Butlerov reaction to the synthesis of any particular sugar. These facts, together with studies of rates of decomposition of ribose and other sugar[17] that show how unstable ribose is, even at neutral pH, present severe obstacles to proposing a prebiotic synthesis of RNA and guide the thought to discard ribose as a component of the first genetic material. Nevertheless, recently it has been reported that pentose sugars are stabilized by borate minerals.[18] The presence of borate is paramount

* A palimpsest is a manuscript that has been erased and used again, but it is still possible to read the older writings.

13. Benner, S. A.; Ellington, A. D.; Tauer, A., *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86 (18)*, 7054-7058.
14. Gaucher, E. A. *et al. Nature* **2003**, *425 (6955)*, 285-288.
15. Decker, P. S. H. P., R., *J. Chromatog.* **1982**, *244*, 281-291.
16. Breslow, R., *Tetrahedron Lett.* **1959**, *21*, 22-26.

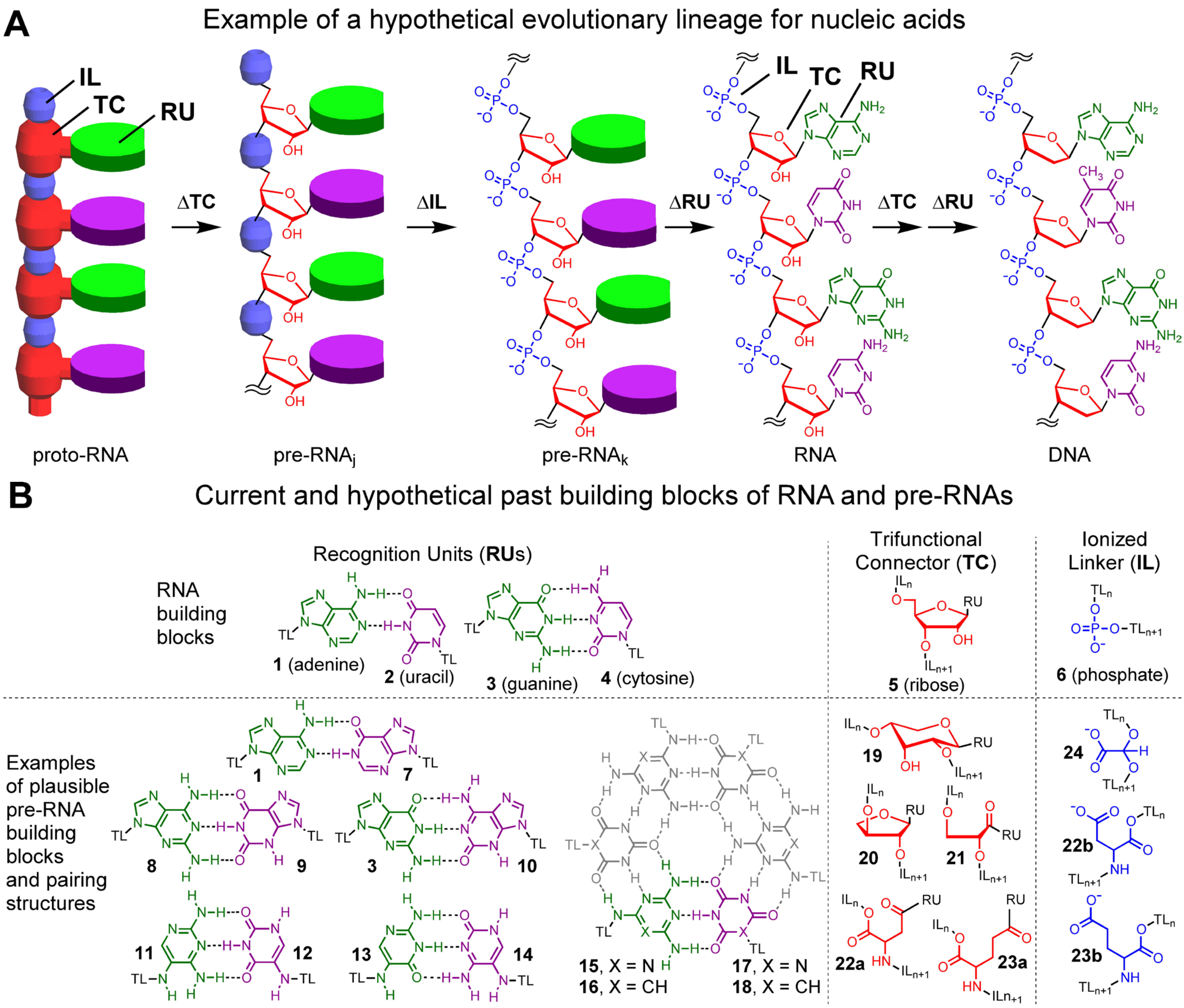


Figure 6. (A) Proposed Evolutionary Pathway to Contemporary Nucleic Acids with Some Plausible Building Blocks of Pre-RNAs 6A Schematic representation of a hypothetical evolutionary lineage of nucleic acids from proto-RNA to RNA and DNA. The three components of RNA are: the recognition units (RUs), trifunctional connector (TC), and ionized linker (IL). Intermediates between proto-RNA and RNA are shown for illustrative purposes only and are not intended to imply that changes in RUs, TC or IL proceeded in the particular order or number of steps shown. (B) Examples of plausible pre-RNA components for RUs, TC, and ILs. Key to structures: 1, adenine; 2, uracil; 3, guanine; 4, cytosine; 5, ribose (furanose form); 6, phosphate; 7, hypoxanthine; 8, 2,6-diaminopurine; 9, xanthine; 10, isoguanine; 11, 2,4,5-triaminopyrimidine; 12, 5-aminouracil; 13, 2,5-diaminopyrimidin-4(3H)-one; 14, 4,5-diaminopyrimidin-2(1H)-one; 15, melamine; 16, 2,4,6-triaminopyrimidine; 17, cyanuric acid; 18, barbituric acid; 19, ribose (pyranose form); 20, threose; 21, glutamine (a as TL; b as IL); 22, aspartate (a as TL; b as IL); 23, glyceric acid; 24, glyoxylate. (Figure kindly provided by N.V. Hud and published in Hud, N. V.; Cafferty, B. J.; Krishnamurthy, R.; Williams, L. D., The Origin of RNA and "My Grandfather's Axe". *Chem. Biol.* **2013**, *20 (4)*, 466-474. DOI: 10.1016/j.chembiol.2013.03.012. Reproduced with permission from the journal. (💻))

for these studies due to the ability to form stable complexes with 1,2-diols. The negative nature of these complexes inhibit the formation of enolates that otherwise lead to tar.[18]

Recently, Nicholas Hud from Georgia Tech and collaborators proposed that RNA is not the product of abiotic processes, but rather is a product of molecular evolution (Figure 6A). Instead of considering RNA as the first polymer capable of storing information, Hud considers RNA as the penultimate member of a series of polymers DNA being the most recent[19] (and the most stable for storing genetic information). Figure 6B shows some possible molecules that could have been part of an older RNA than the one we know. This year, 2014, Hud and collaborators published their work with a one-pot reaction, in prebiotic conditions, where a nucleobases reacts with ribose and assembles with one of the

17. Larralde, R.; Robertson, M. P.; Miller, S. L., *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92 (18)*, 8158-8160. (💻)
18. Ricardo, A. C. *et al*, *Science* **2004**, *303*, 196.
19. Hud, N. V. *et al.*, *Chem. Biol.* **2013**, *20 (4)*, 466-474. (💻)

possible ancestral nucleobases that could have been part of an ancient RNA.[20]

In 2013, the discovery of a ribozyme capable of copying a sequence longer than itself is fueling the hopes for finding a self-replicator.[21] The cellularization of genetic material is the next step and considerable progress in this respect has been done at the Szostak laboratory at the Massachusetts General Hospital (see review by Szostak [22]).

(3) Unusual environments

If we could find life somewhere outside the Earth, this discovery would help us understand the essence of life more than any other discovery, especially if extraterrestrial life would be different than ours at the molecular level. This would imply that life can originate and maintain with different chemical alternatives than the ones we know on Earth. Conversely, if we find living entities with the same biology we know, then we would ask if life originated spontaneously in different places in the cosmos and converged to the system we know or perhaps it was originated in one single place and then transported to different planets (panspermia theory).

We will never know, without a doubt, which was the first self-replicating molecule

The closest candidates for finding life in the solar system appear to be Mars and Europa as these have features (liquid water) that are believed to be essential for life. Europa is one of the satellites of Jupiter, first discovered by Galileo Galilei, and is thought to carry liquid water beneath an ice ocean. Another thought to be considered on this point is the possibility that alien life (meaning living beings with a different biochemistry) is somewhere on Earth and has not been discovered yet. This idea of a shadow biosphere contemplates environments such as hydrothermal vents at the bottom of the ocean, which harbor organisms that thrive at high temperatures and pressure.[23]

(4) Synthetic biology

The understanding of how living organisms work can be defined by the ability of recreating them in the laboratory. Synthetic biology[23] rearranges the molecules of living beings to test/change this product of 3.5 Ga of evolution. After all, if life is nothing more than a chemical system capable of Darwinian evolution, then we should be able to synthesize an artificial system capable of Darwinian evolution. Our laboratories have investigated the effect systematical changes have on the structure of DNA. One of the most important conclusions is that the repetitive charge on the DNA backbone (Figure 2) is fundamental for the recognition of the complementary sequence.[25] The repeating charge might be a universal feature of genetic molecules. This would make us look for a polyanionic molecule as the repository of genes if we are to find alien life.

In the last five years, the field of synthetic biology has made remarkable progress. The laboratory of Phillip Holliger at the MRC Laboratory of Molecular Biology has produced a series of structures analog to the nucleic acids we know and produced enzymes capable of copying the information among these.[27] More recently, the Romesberg laboratory has produced a system in which *E. coli* cells can use genetic alphabet composed of six nucleobases; however, the two extra nucleobases in the Romesberg laboratory do not contain hydrogen bonds.[27]

In our research, we have developed new DNA polymerases able to incorporate nucleobases on DNA with altered hydrogen bonding patterns. This has produced DNA containing two additional letters named Z and P. To do so we have used a system called CSR (compartmentalized self-replication system) in which the genotype and the phenotype are encapsulated in water-in-oil droplets. In this system we can do a selection in the laboratory among millions of variants of polymerase enzymes. These droplets resemble minimalist cells that contain all necessary components needed for amplifying their own gene. In this case the selection barrier is imposed by our experimental design and that requirement is that the new enzymes should be able to incorporate the two extra nucleotides with the same fidelity as they incorporate the four natural ones.

Interestingly, our laboratory evolution experiments using artificial nucleic acids have recapitulated patterns of evolution observed in nature.[28] Remarkably, some of the molecules proposed by Hud (Figure 6B) have the same hydrogen bonding pattern as our artificial pair Z:P. Perhaps one day we will run into a living being that has this extra pair of nucleobases.

20. Chen, M. C.; *et al*, *J. Am. Chem. Soc.* **2014**, *136 (15),* 5640-5646.

21. Attwater, J.; Wochner, A.; Holliger, P., *Nat. Chem.* **2013**, *5 (12)*, 1011-1018.

22. Blain, J. C. and Szostak, J. W., *Annu. Rev. Biochem.* **2014**, *83*, 615-40.

23. Martin, W.; Baross, J.; Kelley, D.; Russell, M. J., *Nature Rev. Microbiol.* **2008**, *6 (11)*, 805-814.

24. Sismour, A. M. and Benner, S. A., *Expert Opin. Biol. Ther.* **2005**, *5 (11)*, 1409-1414.

25. Benner, S. A., Hutter, D., *Acc. Chem. Res.* **2004**, *37 (10)*, 784-797.()

26. Pinheiro, V. B. et al., *Science* **2012**, *336 (6079)*, 341-344.

27. Malyshev, D. A. *et al. Nature* **2014**, *509 (7500)*, 385-388.

28. Laos, R.; Shaw, R. W. and Benner, S. A., "*Engineered DNA Polymerases*". In Murakami, K. S. and Trakselis, M. A., Eds. "*Nucleic Acid Polymerases*", Springer-Verlag, Berlin-Heidelberg 2014; Vol. 30, pp 163-187.

29. Sefah, K. *et al.* *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111 (4)*, 1449-1454.

30. Kratzer, J. T. et al. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111 (10),* 3763-3768. ()

Useful applications

The study of the origin of life might sound like an idealistic crusade that would only be used to satisfy our natural curiosity; however, there are several practical aspects in studying alternative nucleic acids, ancient proteins and the space around us. Besides the numerous technologies developed by NASA in their effort for exploring the cosmos , some of the studies we have mentioned have produced contributions relevant to medicine. Our laboratories, for instance, have developed analogs to nucleic acids that form an Artificially Expanded Genetic Information System (AEGIS). Some of these AEGIS components are currently used in medicine to diagnose patients with AIDS and hepatitis. Nucleic acids containing the additional letters Z and P have been used to detect cancer cells.[29] Additionally, the reconstruction of ancestral enzymes have helped the treatment of diseases like gout.[30]

Outlook

This article reviews some of the landmarks and key information needed to understand the multidisciplinary efforts of the search for the origin of life. The study of the origin of life, once plagued with intractable problems, has now a variety of approaches to find possible solutions. Although we may never know exactly the nature of the last common ancestor, it is possible that in a few years a man-made system capable of Darwinian evolution would be produced and studied in a laboratory. This minimalistic primitive cell would be useful to understand what is the limit between inanimate matter and a living organism. Understanding of the requirements needed to jump from abiotic components present in primitive Earth to a plausible last common ancestor would give us a complete account of the origin of life.

Recent discoveries of hundreds of exo-planets, many of them similar to Earth, are being reported in the news. Nowadays, the question “Where do we come from?” is gaining public interest. Even the Vatican organized an astrobiology conference to consider the possibility of alien life, in November 2009 on the 400th anniversary of Galileo Galilei’s astronomical discoveries. It is clear that the origin of life concerns to everybody.

Acknowledgments

The authors would like to thank Professor Antonio Lazcano for his help in translating some terms from English for the Spanish version of this article. We are grateful to Dr. Nicholas Hud for offering some of his figures for this publication. We are also indebted to support from the NASA under Exobiology/Astrobiology grants NNX14AK37G and RSA1371457.

Received: July 30th 2014
Accepted in final format: September 23rd 2014

* For more information visit: http://spinoff.nasa.gov/ (🖳)

Full spanish version of this paper (translated by the authors) is available here

Essential literature on the subject

Koshland, D. E.: “The seven pillars of life”. *Science* **2002**, *295 (5563)*, 2215-2216. (🖳)

Benner, S. A.: “Understanding nucleic acids using synthetic chemistry”. *Acc. Chem. Res.* **2004**, *37 (10)*, 784-797. (🖳)

Hud, N. V.; Cafferty, B. J.; Krishnamurthy, R.; Williams, L. D., “The Origin of RNA and “My Grandfather’s Axe””. *Chem. & Biol.*, **2013**, *20(4)*, 466-474. (🖳)