

## Major evolutionary transitions before cells: A journey from molecules to organisms

Francisco Prosdocimi <sup>a,\*</sup>, Sávio Torres de Farias <sup>b,c</sup>

<sup>a</sup> Laboratório de Biologia Teórica e de Sistemas, Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

<sup>b</sup> Laboratório de Genética Evolutiva Paulo Leminski, Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba, João Pessoa, Paraíba, Brazil

<sup>c</sup> Network of Researchers on the Chemical Evolution of Life (NorCEL), Leeds, LS7 3RB, UK



### ABSTRACT

Basing on logical assumptions and necessary steps of complexification along biological evolution, we propose here an evolutionary path from molecules to cells presenting four ages and three major transitions. At the first age, the basic biomolecules were formed and become abundant. The first transition happened with the event of a chemical symbiosis between nucleic acids and peptides worlds, which marked the emergence of both life and the process of organic encoding. FUCA, the first living process, was composed of self-replicating RNAs linked to amino acids and capable to catalyze their binding. The second transition, from the age of FUCA to the age of progenotes, involved the duplication and recombination of proto-genomes, leading to specialization in protein production and the exploration of protein to metabolite interactions in the prebiotic soup. Enzymes and metabolic pathways were incorporated into biology from protobiotic reactions that occurred without chemical catalysts, step by step. Then, the fourth age brought origin of organisms and lineages, occurring when specific proteins capable to stack together facilitated the formation of peptidic capsids. LUCA was constituted as a progenote capable to operate the basic metabolic functions of a cell, but still unable to interact with lipid molecules. We present evidence that the evolution of lipid interaction pathways occurred at least twice, with the development of bacterial-like and archaeal-like membranes. Also, data in literature suggest at least two paths for the emergence of DNA biosynthesis, allowing the stabilization of early life strategies in viruses, archaeas and bacterias. Two billion years later, the eukaryotes arouse, and after 1,5 billion years of evolution, they finally learn how to evolve multicellularity via tissue specialization.

### 1. Introduction

According to recent proposals (Farias and Prosdocimi, 2022; Prosdocimi and Farias, 2023), the history of life in Earth can be divided in 4 major ages or eras: (i) prebiotic, (ii) FUCA's, (iii) progenotes', and (iv) cellular organisms' (Fig. 1). This proposition is based on logical assumptions and necessary steps of sophistication along the evolution of biological complexity. Current research in Biology can provide models to explain what happened along the first and the last of those ages, while the two ages in-between are still enigmatic. Along this work, it will be presented an empirically-based, theoretical proposal that describes putative developments and challenges happening along these eras, allowing us to identify the most important events and features that occurred along the early evolution of life in Earth. The current proposal is mainly based on (i) the gradualist assumption that life must have crossed a path from molecules to cellular organisms, and (ii) the understanding about the most important complexification steps that molecules must have passed through in order to evolve into the biological cells we know today. Even if the scenario proposed here is conjectural, we aim to identify necessary steps that allowed the gradual and stepwise

complexification of early molecular biological systems until they reach cellular constitution.

### 2. Prebiotic age, early stage

The most important landmark of experimental simulations about the prebiotic age happened in the 1950's, with the classical experiment of Urey-Miller, on which the renown North American chemist Stanley Miller (1930–2007) made an ingenious experiment to simulate the early Earth atmosphere (Miller, 1953). This experiment inaugurated the field of prebiotic chemistry when simulating a putative atmosphere for the Hadean Earth, which was suggested to contain ammonia ( $\text{NH}_3$ ), hydrogen ( $\text{H}_2$ ), hydrogen sulfide ( $\text{H}_2\text{S}$ ), water ( $\text{H}_2\text{O}$ ) and methane ( $\text{CH}_4$ ). Together with heat and electric discharges, Miller was capable to produce a great number of biomolecules, notably different types of amino acids (Bada and Lazcano, 2002; Parker et al., 2011a; 2011b, 2011c; Bada, 2013). Even if this and further experiments were indeed capable to produce amino acids and small peptides, these entities by themselves would never be capable to evolve into life unless another relevant molecule come into the game: the nucleic acids. However, both the

\* Corresponding author.

E-mail address: [prosdocimi@bioqmed.ufrj.br](mailto:prosdocimi@bioqmed.ufrj.br) (F. Prosdocimi).

environment and the chemical conditions on which the production of their building blocks (nucleotides) were favored are not consensual.

Based on logical assumptions, it is clear that nucleotides must have been produced in prebiotic Earth to allow biological systems to emerge. However, the way on which they have been produced constitute one of the main controversies under debate by the academics studying the origin of life in prebiotic Earth (Powne et al., 2009; Pérez-Villa et al., 2018; Kim and Kim 2019; Becker et al., 2018; Xu et al., 2020). At least four main scenarios have been proposed trying to provide coherent chemical possibilities for the prebiotic production of nucleotides and small nucleic acids, mainly RNAs (Prosdocimi et al., 2022). Some of these scenarios include: (i) the formation of nucleotides in hydrothermal vents found in the deep sea, environments rich in chemical compounds and minerals (Baaske et al., 2007; Helmbrecht et al., 2023). Those chemical compounds provided energy sources and a wide range of chemical reactions that could have facilitated the synthesis of nucleotide precursors. Also, there are chemical scenarios on which nucleotides could be produced in (ii) ponds and lakes on the early Earth's surface. These small watery environments could have provided concentrated amounts of molecules on which complex chemical reactions could occur (Pearce et al., 2017; Ter-Ovanessian et al., 2022). In that case, various energy sources, such as lightning or UV radiation, might have contributed to the formation of nucleotides. It was also proposed (iii) that either nucleotides or their precursors could have been delivered to Earth through the impact of meteorites or cometary materials (Maurette, 1998; Burton et al., 2012; Krishnamurthy et al., 2022). These extraterrestrial sources could have introduced important organic compounds to the planet. Finally, (iv) some works suggested that minerals and clays possessing catalytic properties potentially aided in the synthesis of nucleotides and their precursors by facilitating the chemical reactions necessary to their formation (Nussinov et al., 1997; Pedreira-Segade et al., 2018; Rimola et al., 2019). Those scenarios have been shown to facilitate the production of nucleotides in prebiotic conditions and, although no consensus have been reached about which of them was the most important, we may consider the possibility that all of them happened in parallel (Fig. 2).

### 3. Prebiotic age, late stage

For life to emerge, it is necessary that both small nucleotides and small amino acid polymers have been produced in significant amounts. Now, at the second part of the prebiotic age, the path to life needs that those biomolecules have turned abundant in the prebiotic soup. In terms of logics, only if this happened is that other necessary steps for life to emerge can follow. Thus, at this second-half of Prebiotic age, abundant molecules of replicating RNAs, together with chemical cycles for building metabolites and peptide must have existed. At this point, a sophisticated form of RNA-world may have existed, together with a rich primitive soup on which amino acids, peptides and metabolites cycled according to cosmological and geochemical cycles. It is also under debate by chemists which were those conditions that allowed RNAs to evolve (Prosdocimi et al., 2022). At least medium-sized molecules of RNAs should have existed at that time, possibly harboring dozens of nucleotides. Those molecules probably folded in some nearly stable tRNA-like shape and were capable to bind other molecules in the pool,

and self-replicate (de Farias et al., 2016). At that time, the binding of free and abundant nucleotide molecules to RNA polymers should have happened. These stacking of nucleotides in the RNA polymers would modify their function, structure and stability, and the stacking of many nucleotides would allow template-based replication (Szilágyi et al., 2017; Vörös et al., 2021). At that time, it is possible to imagine that RNA molecules could grow in two ways: (i) by acquiring new nucleotides, probably in the 3' terminus, and (ii) by allowing free nucleotides to bind and performing a plus-minus replication of a complementary, anti-parallel strand. The modification of RNA polymers by the pairing of single nucleotides would change their folding, allowing different interactions and different stabilities. Those modifications will evolve into the binding of nucleotidic-based cofactors to RNAs and peptides (Kirschning, 2021).

The physicochemical and cosmological cycles in early Earth, such as day-night and seasonal gradients allowed some conditions to happen again day by day, season by season, and year by year. Initially, simple molecules were formed but then, when they became abundant, more complex molecules and reactions could happen, allowing a chemical complexification of the prebiotic soup. Randomly formed small peptides would fold and bind metabolites present in the soup, therefore acquiring different structures and affinities (Prosdocimi et al., 2021). As the chemical complexity got higher, the auto-organization process evolved by walking locally against the thermodynamics gradient, lowering the entropy and allowing the emergence of auto-organization forces. On the other hand, as many protobiological molecules got more complex through the chemical symbiosis by (i) binding other molecules, (ii) becoming more stable and (iii) triggering autorganization processes, a much greater number of other molecules would fall into chaos and extinction.

Several models explored the formation of self-referential cycles for establishing the initial interactions in early biological systems. Although their details diverge, the organizational principles follow a similar logic, on which catalysts and products connect to stabilize and bring resilience to the system (Cornish-Bowden and Cárdenas, 2020). In this context, the relationship between catalysts and products expands into various nodes of a self-sustaining network, forming a hypercycle that can be understood as an autopoietic process. Those systems develop networks of interactions capable to recreate the network itself. Their growth occurs through the incorporation of new elements that do not disturb neither stability nor robustness (Maturana and Varela, 1973; Eigen and Schuster, 1977).

Thus, the chemical organization in the prebiotic age followed two essential paths: (i) the formation and evolution of small RNAs and the RNA-world, and (ii) the evolution of quasi-random peptides capable to bind, stabilize and cycle together with metabolites in the prebiotic soup. Those separated chemical worlds probably evolved independently and existed in different prebiotic refugia in early-Earth (Prosdocimi et al., 2022). At that point, these systems functioned as a set of autocatalytic processes (Kauffman, 1986), on which autocatalytic structures emerged without the influence of selective processes. We should also consider that the self-organization of these isolated populations were target by independent mutations brought by genetic drift (Dyson, 1982). At some point, those worlds of self-catalyzing nucleic acids and peptides hypercycles will meet through an event of chemical symbiosis, which will

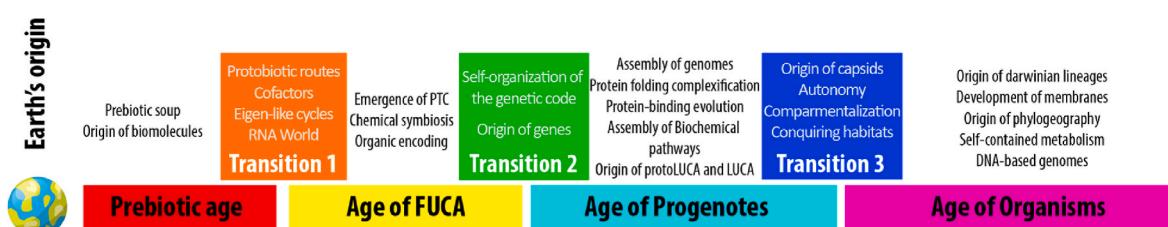
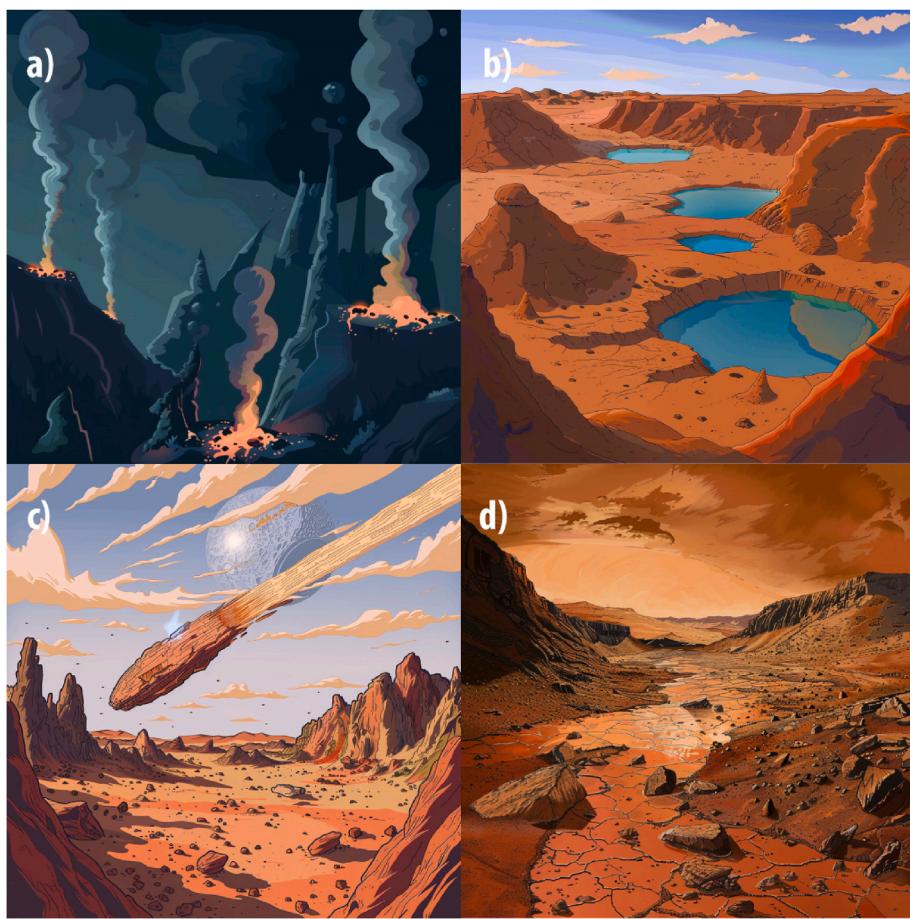


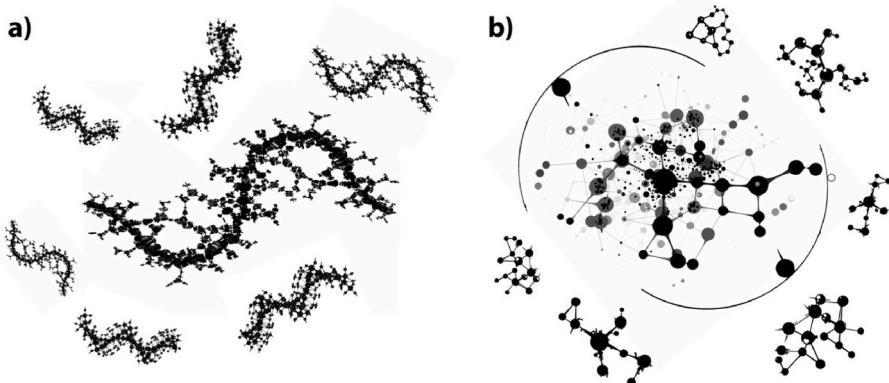
Fig. 1. The four ages of life in Earth and their major transitions.



**Fig. 2.** The four major scenarios described in literature for nucleotide production in the prebiotic age operating in primitive Earth: (a) hydrothermal vents; (b) lakes and ponds; (c) meteorite origin; (d) clay mines and veins.

allow them to step forward into a significant higher level of auto-organization that will inaugurate biology (Prosdocimi et al., 2019, 2021). The integration between the self-referential cycles established by nucleic acids and peptides thus enabled the organization of new properties in the emerging system. This allowed for the emergence of a process of correspondences between these different classes of chemical polymers, thereby creating interdependence and inaugurating the establishment of the encoded biological information. Such a process of

correspondence and dependency occurred through the initiation of coding process, allowing the maintenance of a self-referential system and inaugurating the biological systems as processes based on chemical translation of codes (Farias et al., 2021). At this point, nascent biological systems transitioned from a self-organized process guided by drift to a model guided by the process of natural (molecular) selection, where the polymers that established self-reference relationship with others could be continuously recreated and maintained within the system.



**Fig. 3.** The emergence of life was catalyzed by a chemical symbiosis between two previously independent realms of molecules, depicted schematically as follows: (a) the RNA world, consisting of self-replicating nucleic acid molecules, and (b) the peptide-metabolic world, on which physicochemical cycles generated metabolites within the prebiotic soup. Of particular interest are the cycles capable of synthesizing amino acids. These two domains of macromolecules became integrated through a chemical symbiosis event. In this contingent event, an RNA molecule folded into a specific 3D structure capable to bind amino acids together, thereby giving birth to FUCA as the proto-PTC.

#### 4. The age of FUCA, early stage

It was only when an enriched prebiotic soup containing amino acids, small peptides and metabolites became chemically connected to the self-replicative RNA molecules that the principles governing the life phenomenon as we know it could emerge (Prosdocimi et al., 2021). Neither the RNA-world nor the self-catalytic hypercycles of metabolites by themselves should be considered alive according to this reasoning (Fig. 3). Thus, early prebiotic systems will find their path to life when a chemical symbiosis between nucleic acids and peptides auto-assemble a code-based, stable relationship (Prosdocimi and Farias, 2019; Farias et al., 2021).

Recent theories suggest that this connection between (i) the nucleotide and (i') peptide/metabolite worlds happened when a preliminary version of the ribosome arose (Dantas et al., 2021). At some point, biologists described the contingent evolution of an ancient nano-machine made of an RNA molecule capable to bind amino acids together (Davidovich et al., 2010). This nano-machine was capable to form peptide bonds, allowing the formation of quasi-random small peptides (Davidovich et al., 2009). At that point, the composition of peptides depended on the availability of certain types of amino acids in some specific prebiotic refugee (Prosdocimi et al., 2022). This nano-machine started its self-assembly by the emergence of the preliminary version of the peptidyl transferase center (PTC), that is the ribozyme responsible to bind amino acids in current ribosomes (Yonath et al., 2000). By means of self-organization and accretion, these proto-PTCs, formed by self-replicative RNAs capable to bind amino acids covalently, allowed the most significant event in early life to occur: the chemical symbiosis (Prosdocimi et al., 2021). Being a sort of molecular mutualism (Lanier et al., 2017; Vitas and Dobovišek, 2018), this interaction was beneficial to both RNAs and peptides, allowing them to grow in size and complexity, as well as increasing their stability. This event will follow the formation of an initial form of a proto-ribosome, followed by a more sophisticated level of self-organization: the emergence of the first biological code (Barbieri, 2003, 2014; Prosdocimi and Farias, 2021, 2023). Prosdocimi and collaborators (2019) named these early populations of self-replicating, peptide making RNA entities as the First Universal Common Ancestor (FUCA). According to these authors, FUCA was (i) born when an early form of the proto-PTC emerged and (ii) matured when an early (though functional) version of the genetic code was established (Prosdocimi et al., 2019).

FUCA should therefore be considered as the first biological system that inaugurated an era of self-reproducing populations of ribonucleoproteins. At the center of this complex lies an RNA capable to bind amino acids covalently, reproduce and complexify, giving rise to more complex molecular subsystems (Prosdocimi and Farias, 2023). Thus, the ultimate essence of life is a ribonucleoprotein process of chemical encoding and information exchange (Farias and Prosdocimi, 2022). Then, different subsystems will descend from this original population of FUCA, that may have emerged only once. These ribonucleoprotein agglomerates will become each time more stable, organized and complex, stepping against the physicochemical annihilation process of entropy. We may consider that millions of other RNP-like entities have been annihilated while only a few specific ones could climb the stairs of complexification. Thus, at the first stage of FUCA's age, it occurred the stabilization of the encoding and replicative processes of those RNP entities. Once these issues were well resolved, FUCA could reproduce better and be the ancestor of populations of descendants. Those descendants of FUCA were expected to be displaced from their center of origin and migrate along other chemical refugia at the primitive soup environment, being therefore subjects of different mutations and patterns of evolution.

#### 5. The age of FUCA: folding and stability of small RNAs

The most stable structural fold for small RNAs is the stem-loop or hairpin structures. These structures are characterized by short regions of

double-stranded RNA (the stem) connected by a single-stranded RNA loops. The stem of the hairpin structure was formed by base-pairing between complementary nucleotides, allowing stability and structural rigidity. On the other hand, the loop region contained unpaired, single-stranded nucleotides that provided flexibility and allowed those polymers to interact with other molecules in the prebiotic soup, such as other amino acid, metabolites and ions. Thus, FUCA entities were formed by RNP molecules, on which simple, self-catalytic, amino acid binding RNAs formed some sort of chemically-stable agglomerates with peptides (Farias and Prosdocimi, 2022). Since its birth, FUCA formed therefore some earlier form of a ribonucleoprotein chromatin that contained both the information exchange core but also some sort of auxiliary RNPs helping in creating a protection and pseudo-compartmentalization. This sort of RNP pseudo-compartmentalization is observed nowadays inside cells and has been known as liquid-liquid phase separation (Koga et al., 2011; Gomes and Shorter, 2019; Zhao and Zhang, 2020). We suggest that these events happened in a simpler form since the emergence and maturation of FUCA-like entities (Prosdocimi and Farias, 2023). Plus, we suggest that those surrounding RNP particles responsible for stabilization of the agglomerates were probably produced by small molecules that resulted by incomplete replications of the informational core and truncated peptides of the first genes.

The stability of the first stem-loop structures occurred due to the formation of hydrogen bonds between complementary nucleotides, which provide structural integrity. Additionally, stacking interactions between adjacent base pairs and other stabilizing forces, such as base stacking and base pairing mismatches, contributed to the structural stability of RNAs (Stephenson et al., 2013). It is important to notice that while the stem-loop is generally considered the most stable structure for small RNAs, other secondary structures, such as pseudoknots or complex tertiary interactions, can also be formed in certain cases and contribute to higher the RNA stability and allow functional motifs to operate. Also, we must consider the relevance of small peptides to assist and stabilized those structures, always considering RNP agglomerates. Thus, the specific structural folds adopted by small RNA molecules bound to amino acids will depend not only on their sequences and structures, but also in their length and the conditions found in the surrounding chemical environment at that specific refugee in the primitive soup. Nowadays, hairpin structures are commonly found in various functional RNA molecules existing inside cells, including microRNAs, small interfering RNAs (siRNAs), transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs). These molecules still interact with peptides and proteins, playing essential roles in processes such as gene regulation, RNA interference, RNA and protein stabilization and half-life, and, of course, in protein synthesis. However, some of those genetic processes operated quite differently at the time on which the first genomes were produced.

#### 6. The age of FUCA: self-organization of the proto-translation system

Using statistical models for reconstructing the ancestral state of nucleotide sequences, De Farias (2013) downloaded all the 20 current families of tRNAs responsible to bind different amino acids found in Bacteria and Archaea, and produced ancestral sequences for each, naming them as proto-tRNAs. As expected, those 20 ancestral proto-tRNAs were shown to be smaller than the current ones present in life forms, once early molecules are expected to be simpler. In an interesting bioinformatics experiment, this author concatenated different proto-tRNAs families and BLASTed those concatamers against current protein databases using BLASTx (Altschul et al., 1990). De Farias (2013) found that some of those concatamers presented similarity to enzymes involved in biochemical pathways often described as fundamental, such as proteins forming the glycolytic and citric acid pathways. More importantly, he found that a sequence generated by the concatenation of a specific set of proto-tRNAs (Proline, Tyrosine, Phenylalanine, Glutamine and Glycine) presented more than 50% of sequence identity

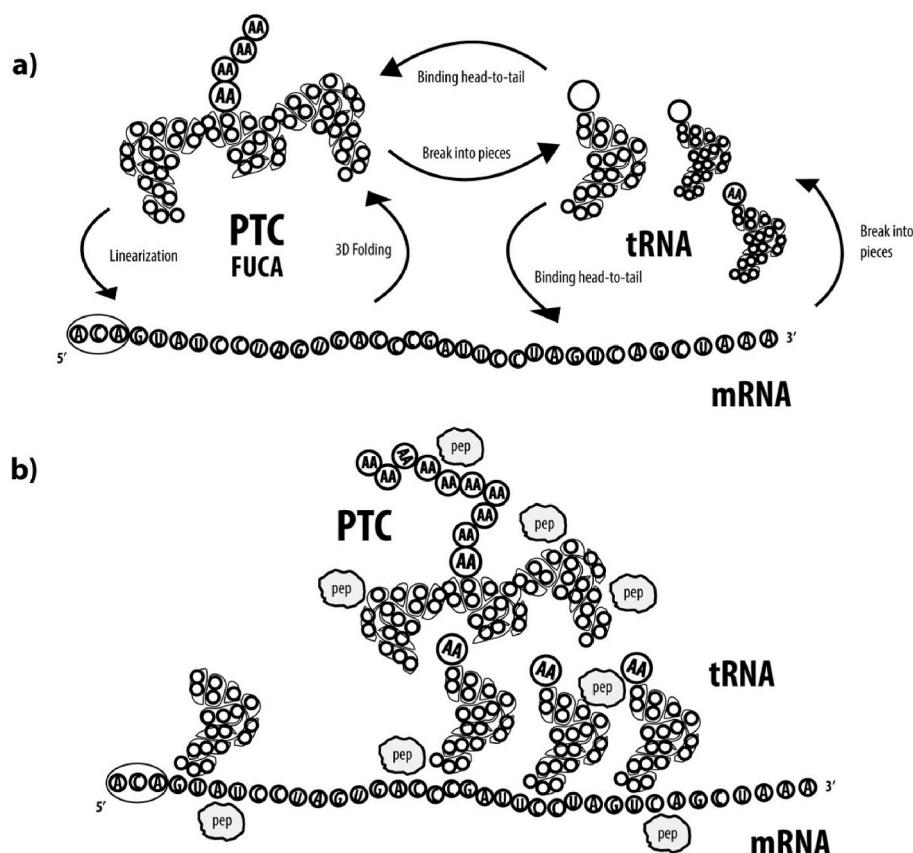
to the PTC of the living bacteria *Thermus thermophilus* (De Farias, 2013). In a further work, a research group demonstrated that this putative form of a proto-PTC presented 92% of structural similarity to the current PTC of *T. thermophilus* (de Farias et al., 2017). Inspired by this work, Prosdocimi and collaborators (2020) demonstrated that this early, concatenated form of a proto-PTC was capable to fold into two different structures, one allowing the formation of (i) the PTC ribozyme, and the other (i') facilitating the self-replication of this molecule (Prosdocimi et al., 2020).

Root-Bernstein and Root-Bernstein (2015) found similar results analyzing ribosomes of bacteria, demonstrating that the modern rRNAs present several molecule markers of tRNAs in their sequences. Thus, they proposed not only that (i) the primitive ribosome was formed by a concatemer of tRNAs, but also that (ii) rRNAs probably functioned as primordial genomes, being possibly encountered in linear structures and forming primitive forms of mRNAs (Fig. 4). In addition to these data, several groups have obtained results suggesting that the PTC was the first part to be structured in the ribosome; and that the ribosome became more complex by the addition of new parts over the PTC region (Petrov et al., 2014; Petrov and Williams, 2015). Recently, Bose et al. (2022) synthesized molecules corresponding to the primitive ribosome, containing only parts of the PTC, and demonstrated that these primitive ribosomes could indeed form peptide bonds randomly.

Corroborating this idea, it has been demonstrated that the smaller subunit of the ribosome also exhibits similarity, both in terms of primary sequence and three-dimensional structure, with ancestral tRNA molecules (Farias et al., 2021). In another publication, de Farias et al. (2017)

provided evidence that the ancestral PTC could interact with those proto-tRNAs, this interaction being similar to the one observed in modern ribosomes. Thus, tRNAs may have functioned as a delivery system for amino acids to be polymerized by the primitive PTC. Advancing in the understand about the mechanistics of genetic code self-organization, Farias et al. (2021) suggested that the interaction process between the proto-PTC and the primitive decoding center placed in the smaller ribosomal subunit created a restriction system to allow the decoding process. This way, only tRNAs carrying an anticodon loop complementary to the proto-mRNA (and therefore bringing their specific “encoded” amino acids) would be able to bind the proto-ribosome.

Thus, a key issue in the emergence of the genetic code that allowed FUCA to mature is the understanding about how the code was self-organized. Divergent models try to provide a mechanistic, chemical-driven theory to explain why some codons are associated to specific amino acids. One model suggests that the genetic code structured itself through random events and became frozen when physicochemical forces and error patterns were minimized (Freeland and Hurst, 1998). Others suggest that the observed correlations are the result of a chemical affinity between the triplets and their respective amino acids (Farias et al., 2007; Guimarães et al., 2008). Also, alternative models suggest that the code structured itself in parallel with the development of metabolic pathways for amino acid synthesis (Davis, 2002; Xue et al., 2003). Although initially these propositions seem self-excluding, they can actually complement each other if we assume that multiple forces were involved in the genetic code's emergence.



**Fig. 4.** The origin of genetic decoding process probably happened with RNA molecules with similar sequences but diverse functionalities dictated by slight variations in their 3D conformations. (a) Small tRNAs exhibited dual capabilities: (i) covalently binding amino acids at their 3' ends, and (ii) linking together head-to-tail to form the proto-peptidyl transferase center (proto-PTC). This proto-PTC possessed the ability to bind amino acids, catalyzing the formation of peptides. Additionally, the proto-PTC might also assume linear configurations resembling proto-mRNA molecules. These proto-mRNA molecules not only facilitated nucleotide stacking for self-replication but also orchestrated the sequential arrangement of tRNAs. (b) The primitive synthesis of proteins in FUCA entities involved the interplay between specialized RNAs and small stabilizing peptides (depicted as pep).

## 7. The age of FUCA: the evolution of aminoacyl tRNA synthetases

To investigate this issue further, we must consider the origin and evolution of aminoacyl tRNA synthetases (AAS) and the molecular selection operating in their specific binding to certain tRNAs and amino acids. Defining the fidelity of biological information encoding, AAS present two sub-families, each containing 10 members and responsible for the aminoacylation of an amino acid to its corresponding tRNAs. These classes show no significant sequence or structural similarities, suggesting independent origins (Carter Jr, 2017; José et al., 2023a, 2023b). In modern organisms, tRNAs bind amino acids at their 3' end, known as the acceptor arm. This is the only known case on which a nucleotide of adenine performs a covalent linkage to an amino acid, and the binding is performed between the ribose's 3'-OH and the carboxylic acid of the encoded amino acid. As the amino acid binds the acceptor arm at one side of the tRNA, the decoding site of the mRNA occurs at a loop of unpaired bases located at the other structural side of the tRNA, the anticodon loop. These two portions of tRNAs are separated by a distance of ~70 Å, and it has been demonstrated that the acceptor arm can be aminoacylated without the presence of the anticodon loop (Dantas et al., 2021). This corroborates the idea of an operational code that originally functioned without codification (provided by the anticodon loop) (Rodin and Ohno, 1997; Shore et al., 2020). Interestingly, another study demonstrated that the anticodon loop could also be aminoacylated without the presence of the acceptor arm (Dantas et al., 2021).

Farias and colleagues (2008) demonstrated that there is a hydrophobicity correlation between some amino acids and the anticodon of tRNAs. Dantas et al. (2021) reconstructed the sequence and ancestral structure of all class I AAS and analyzed their binding capacity. Their data suggested that both the anticodon loop and the acceptor arm exhibited affinities for the catalytic site. Thus, they propose that the interaction of the anticodon loop shifted to another portion of the protein, while the hydrophobicity relationships were maintained. However, these studies explain only start to explain how the encoded information was self-organized. Before the emergence of FUCA, it is clear that proto-tRNAs bound with more affinity to certain amino acids due to physicochemical forces alone, and these bindings allowed the reading of the information contained in primitive mRNAs. Also, some studies demonstrated that certain ribozymes could aminoacylate tRNAs before the origin of aminoacyl tRNA synthetases (Xu et al., 2014; Janzen et al., 2022). Therefore, the emergence of the encoding process may have been initiated with ribozymes and later replaced by peptides.

Regarding the origin of the first mRNAs, Farias et al. (2014) analyzed the coding capacity of ancestral tRNAs when concatenated. They observed similarity with modern proteins involved in translation and aminoacyl tRNA synthetases. Thus, de Farias et al. (2016) proposed the tRNA-core hypothesis, suggesting that primitive tRNAs orchestrated the maturation of FUCA by self-organizing the formation of primitive translation systems and the origin of the first genes. Finally, at the end of FUCA's age, FUCA-like RNP entities with a functioning translation system self-replicated and expanded in the environment, giving rise to a new sort RNP entities: the progenotes.

## 8. The age of progenotes

The age of progenotes inaugurated a new era in the evolution of life in Earth. The original concept of progenotes, as described by Carl Woese, is ideal for characterizing these proto-organisms formed by living, self-replicating RNP particles capable to process an organic code and evolve (Woese, 1998). The realm of progenotes will persist throughout the entire transition from FUCA to LUCA (Prosdocimi and Farias, 2020) and it is the most unknown age on the evolution of life since few works tried to conceptualize and study this age. The central question is the understanding about how self-replicating RNP particles evolve into

cellular constitution (Prosdocimi and Farias, 2023). This is a long non-told story we aim to provide some logic-driven clues and conjectures here.

The study about the early evolution of progenotes must consider two crucial aspects: (i) how FUCA-like RNP entities started to accumulate RNA and peptides in the form of long genomes bound to peptides in some sort of proto-chromatin structures, and (ii) how the biochemical pathways originated and were further united together in a single organism like the LUCA (Farias and Prosdocimi, 2022) (Fig. 5). The events described below will provide necessary paths for the spread and diversification of progenotes, a sort of free-living RNP complexes capable to process organic codes that will evolve to seed two new strategies for life maintenance: viruses and cells (de Farias et al., 2019).

## 9. The age of progenotes: the evolution of early RNA genomes and proto-chromatin

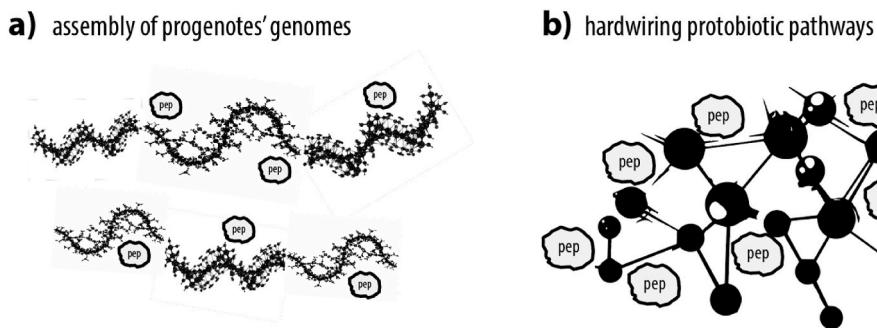
Therefore, we envision a scenario on which the formation of the FUCA and the proto-ribosome happened earlier than the evolution and growing of those RNA genomes associated by proteins that would function a gene repositories (Prosdocimi and Farias, 2023). There would be no genes without genetic codification and the processing of chemical encoding information. Therefore, this order is logical and coherent. We saw that the early ribosomal RNA that formed the PTC probably acted also as the first messenger RNA (Root-Bernstein and Root-Bernstein, 2015), being capable to translate a copy of itself and generate quasi-random peptides at the origin of a proto-PTC. As soon as these first peptides were formed and the genetic code has been better established, we can imagine these free-living progenotes were capable to translate not only the information encoded in their RNAs but also the information coming from other RNAs available in the soup (Prosdocimi and Farias, 2023). Although we will be describing in this section the evolution of RNA genomes, we must always consider that these molecules were constantly bound to amino acids and peptides that formed a primitive sort of chromatin structure. Therefore, although we will be talking about RNAs genomes, please consider that those growing pieces of RNAs were involved by stabilizing peptides both bound directly to them but also helping to aggregate their surroundings by forming a sort of RNP colloid media that enhanced their stability.

The presence of specific RNA populations in an environment was governed by molecular natural selection. This means that RNAs capable to: (i) endure more, (ii) self-replicate faster, (iii) present more fidelity in their copies, (iv) connect better with the proto-ribosomes and (v) produce peptides that would help them in these issues, would be selected. Natural selection operating in the molecular level would favor different populations of RNAs presenting those characteristics. Different computational models have tried to understand how those early RNA genomes could replicate and evolve (Takeuchi and Hogeweg, 2012; Mizuuchi et al., 2022). Of course, different populations will have one or other characteristics more prominent than others, allowing natural selection to happen in the molecular level. At some point, those molecules probably found some way to bind one-another in a head-to-tail fashion, allowing them to grow and produce the first RNA-based genomes associated to proto-chromatin-like peptides.

## 10. The age of progenotes: concatenation and recombination of RNAs

It is well-known that RNAs have the capacity to grow by concatenation (Erlenbach et al., 2019) through a process called template-independent ligation (Kuhn and Frank-Kamenetskii, 2005) or, simply, RNA concatenation. This process involves the joining of two or more RNA molecules together, resulting in the formation of longer RNA molecules (Stark et al., 2006; Stark and Rader, 2014).

Nowadays, *in vitro* molecular biology experiments, template-independent ligation of RNAs can only occur under certain conditions,



**Fig. 5.** The two major developments along the age of progenotes: (a) the assembly of genomes through the concatenation of genetic material from different progenotes' populations, and (b) the hardwiring of probiotic pathways. This occurred as randomly formed peptides encoded in progenote genomes (depicted by pep) happened to bind metabolites from the primitive soup (represented by black balls) and expedited their cycling, serving as catalysts.

in the presence of appropriate enzymes or chemical catalysts (Malagoda Pathirana and Martin, 2023). One of such enzymes that facilitates RNA concatenation is the T4 RNA ligase, which is commonly used for RNA manipulation in laboratory (Violette et al., 2011; Zhuang et al., 2012). The current processes of RNA concatenation are often described in 4 consecutive steps: activation, joining, annealing, and ligation. In the (i) activation process, RNA molecules are typically phosphorylated at their 5' ends; then the (ii) joining happens when two or more RNA molecules are mixed together in the presence of the appropriate ligase enzyme and reaction buffers. The ligase catalyzes the formation of phosphodiester bonds between the 3' end of an RNA molecule and the 5' end of another, leading to their covalent bonding. This event is often followed by (iii) annealing, where complementary sequences between the different RNA molecules align and form base pairs, providing stability to the RNA molecules to be concatenated. Finally, (iv) at the ligation process, the ligase enzyme continues the process of binding by joining additional RNA molecules to the growing concatemer. RNA concatenation has been used in various experimental techniques and technologies, including the construction of long RNA templates for *in vitro* transcription, the generation of RNA libraries, and the production of RNA-based materials or nanodevices (Simmel, 2007; Yu et al., 2021).

It is important to note however that, in natural biological systems, RNA concatenation is not a common occurrence. RNA molecules are nowadays produced individually through the transcription from DNA double-stranded templates. Although the concatenation of RNA molecules is a laboratorial technique applied today within the realm of experimental molecular biology focused on RNA research, we aim to suggest that similar steps might have happened in prebiotic conditions (Chambers and Patrick, 2015). In that case, ribozymes and other cofactors may have replaced the action of the ligase enzymes used in the laboratory nowadays (Lei and Burton, 2023). This process of RNA concatenation probably happened in the primordial soup and allowed that different populations of evolving RNAs could grow and be mixed, sharing their properties, and evolving longer self-replicative RNA populations (DasGupta et al., 2023). So, the emergence of larger RNA genomes became possible.

### 11. The age of progenotes: the emergence of metabolic pathways

Some hypotheses suggest that certain fundamental chemical reactions involved in early metabolism may have existed before the emergence of biological catalysis (Keller et al., 2014). Prosdocimi and Farias (2022) argue that these metabolic exchanges operating independently of any biological catalysts played a crucial role in guiding the origin of metabolism and named them as "probioptic pathways". These set of reactions were driven by naturally occurring physical and chemical processes in the primordial soup environment, such as cosmological cycles, geochemical reactions or energy provided from external sources

like heat, radiation, or electrical discharges. Under specific conditions, basic chemical transformations and reactions occurred and lead to the formation of simple organic molecules (Prosdocimi et al., 2022).

Recent discoveries by a research group led by the geneticist Markus Ralser at the University of Cambridge have shed more light on this topic (Keller et al., 2014). By analyzing the composition of sediments in oceans, they reconstructed the hypothetical chemical metabolites present in Earth's oceans before the origin of life. Their findings provided evidence for the prebiotic existence of numerous metabolites found in standard biochemical pathways (Luisi, 2014). Consequently, new hypotheses regarding the origin of metabolism have emerged, suggesting that the reactions comprising the fundamental metabolic pathways predate the emergence of biological catalysis (Ralser, 2018; Muchowska et al., 2019a, 2019b; Xavier et al., 1922). These studies indicate that some form of abiotic cycling of molecular transformations occurred in prebiotic environments guided solely by the laws of physics and chemistry. Therefore, it is possible that certain portions of the metabolic pathways were already established prior to the emergence of biological systems (Prosdocimi and Farias, 2022). Subsequently, as more efficient biological catalysts evolved, they were incorporated into these pre-existing, probioptic cycles of molecular information, either by replacing or by modifying the existing physicochemical flow of metabolites (Ralser, 2018). In this regard, we proposed that enzymes did not invent entirely new reactions; rather, after the quasi-random appearance of proto-enzymes in the progenotes, they co-opted already existing reactions on the prebiotic Earth, facilitating the occurrence of those same reactions or similar ones by reducing their cycling time through catalysis (Prosdocimi and Farias, 2022). Interestingly, these findings can be seen as a reinterpretation of the classic concept of hypercycles proposed by the German chemist Manfred Eigen (Eigen, 1971; Eigen and Schuster, 1977; Eigen et al., 1980).

Peptides are known for their expertise in molecular binding, thanks to the significant chemical diversity achieved through the sequential arrangement of amino acids with diverse physicochemical properties. Moreover, peptides exhibit remarkable 3D structural diversity owing to their folding capabilities. Thus, we can envision quasi-random peptides generated during the development of the protein synthesis apparatus that had the ability to bind to molecules present in the prebiotic soup involved in probioptic pathways. If such binding enhanced or accelerated the probioptic reactions and increased the fitness of progenote populations, the production of these peptides would be favored by molecular natural selection (Prosdocimi and Farias, 2022) (Fig. 5b). Consequently, progenotes evolution based on their interaction with preexisting molecular pathways working without biological catalysts. As encoded peptides bound to metabolites in prebiotic soup, natural selection acted to maintain those peptides capable of increasing the speed or stability of some progenote populations. This process illustrates how physicochemical reactions evolved towards the utilization of biological catalysts. Thus, the entire protein metabolism has been "inspired" and

co-opted from chemical reactions occurring in the protobiotic environment (Prosdocimi and Farias, 2023). A mechanism that can explain the evolution of those paths are based in the quasi-random appearance of encoded peptides that were capable to bound metabolites in the prebiotic soup. Those peptides and their encoding RNA populations assisted protobiotic pathways and evolved through mutation, drift, and natural selection. This allowed those systems to became more complex, ultimately giving rise to the biochemical pathways found in the genome of the LUCA (Prosdocimi and Farias, 2023). In order to produce a complex progenote like LUCA, containing dozens of biochemical pathways and hundreds of gene families (Weiss et al., 2016), it has been necessary that those paths should develop in different progenote populations and later shared through the process of lateral transference and genome concatenation (Prosdocimi and Farias, 2023).

Several authors proposed that the maintenance of an autopoietic or self-referential state could only be possible when the whole system presented boundaries and their cycles and reactions are occurring within this semi-permeable space (Maturana and Varela, 1973; Ganti, 2003). In this sense, Lai and collaborators (2021) demonstrated an increase in the catalytic efficacy of ribozymes when they are enclosed by membrane-like structures (Lai et al., 2021). Igamberdiev and Kleczkowski, (2023) highlighted the importance of an enclosed system for the establishment of a robust autopoietic process. In their model, these researchers suggested that the establishment and stabilization of metabolism, with reactions predominantly performed by enzymes, probably occurred after the emergence of the boundaries that individualized nascent biological systems (Igamberdiev and Kleczkowski, 2023).

Although our model, at first glance, appears to conflict with these models that propose a closed system for the stabilization of metabolism, we suggest that this stage of metabolic structuring in an era of progenotes represented a preparatory phase for the establishment of autopoietic or self-referential systems. In this context, we can think that the evolutionary path to the encapsulation of progenote populations occurred precisely due to fact that encapsulation increased the robustness of the systems, being an essential evolutionary force for the transition from self-organized open systems to closed systems, inaugurating a new era in the history of life on the planet: the era of organisms. In any case, our model suggests that ribonucleoprotein aggregates were capable of forming types of colloids that separated even at a liquid-liquid interface without a lipid membrane. These same aggregates can still be found in certain biological systems today microenvironments inside cells (Koga et al., 2011; Gomes and Shorter, 2019; Zhao and Zhang, 2020).

## 12. The age of organisms, early stage: the emergence of viruses

The origin of viruses has been very much disputed recently since a new interest emerged in virology after the COVID-19 pandemics. Also, with the discovery of giant viruses, a great era of controversies about the evolution of virus began, with some researchers arguing in favor of considering viruses as a 4th domain of life (Colson et al., 2012), while some suggesting a polyphyletic evolution of viruses (Koonin and Yutin, 2010; Bäckström et al., 2019) and others proposing that viruses probably evolved from pre-existing cells, maybe from an extinct domain of life (Colson et al., 2018; Barreat and Katzourakis, 2023).

We found more convincing the evidence that support the polyphyletic evolution of viruses. Viruses compose different strategies of life based on enveloped RNPs capable to process the organic code and produce proteins (de Farias et al., 2019). Although it is clear that many viral clades evolved from cells that lost genetic content, two main lines of evidence points us to the fact that viruses probably evolved before cells. The first lies on the argument of (i) simplicity, as the viral organization is much simpler than the cellular one. Besides, viruses are simply formed by RNP particles, the most basic units on which life is based (Farias and Prosdocimi, 2022). Second, there is the argument of (ii) abundance, as many estimations suggest that there are more viruses

in Earth than stars in the whole universe. The number of viruses estimated to exist in Earth reaches the amazing level of  $10^{31}$ , exceeding largely the amount of stars in the universe (Suttle, 2013).

To consider criticisms, most researchers that defend the late origin of virus argue in favor of the fact that viral entities are not autonomous, *i.e.*, they need cells to replicate. Our approach suggest that both virus and cells are different strategies that evolved in the process of self-organization of live in Earth (de Farias et al., 2019). We consider both virus and cells as polyphyletic, therefore being originated multiple times (Prosdocimi and Farias, 2023). At the time of progenotes, we can imagine some progenote populations on which a protein capable to stack together with other subunits of itself appeared. These proteins would allow the capsids to start structuring, further evolving into proto-capsids. Corroborating the argument of simplicity, a great number of current living viral clades present capsids formed by a single protein that stacks together with its copies forming “icosahedral capsids”. These clades compose a significant number of viral families, including: Picornaviridae, Adenoviridae, Herpesviridae, Papillomaviridae, Parvoviridae and Reoviridae, among others (Rossmann and Johnson, 1989; Almendral, 2013; Mietzsch et al., 2019). Thus, the origin of one single gene could be responsible for this great evolutionary achievement. For a progenote, this proto-capsid would allow extra protection and may have permitted progenotes to cross microenvironments they could not conquer earlier (Farias and Prosdocimi, 2022). The emergence of capsid-like structures allowed progenotes to conquer new environments and prebiotic refugia (Prosdocimi et al., 2022), seeding life along the planet.

The maturation of protein synthesis together with its structural stabilization, as well as the structuring and diversification of basal metabolic pathways, still in a semi-open context, allowed the formation of complex populations of progenotes. Thus, the stabilization of protein structures enabled the aggregation of those protein complexes now known as capsids around certain progenotes' populations. This compartmentalization of progenotes could occur with or without the presence of the translation system. If it occurred without it, there would be a need to maintain access to the proto-ribosome, a crucial issue for both for the replication of genetic material and the synthesis of structural proteins (de Farias et al., 2019). At this point, the first group of non-autonomous organisms emerged, as they began to require FUCA-like translator entities for the maintenance of their activities.

Therefore, organisms should be understood as evolutionary products of the co-option process of metabolic groups or networks during the compartmentalization process. Those entities will become non-autonomous when the translation system is not incorporated along this evolutionary path, or they will be autonomous when the translation system is incorporated along the compartmentalization. Thus, from a historical perspective, non-autonomous organisms like viruses never needed cells for their replication, but they simply need access to a ribosome to allow their replication and gene expression (de Farias et al., 2019). This explains why viruses are typically considered cell-dependent when they are merely dependent on a translation system. Although nowadays these FUCA-like protein translation systems are only found inside cells, this was not the case during the time of progenotes, on which ribonucleoprotein FUCA-like subsystems existed and replicated in the primordial soup, assisted by RNP particles in their surroundings.

We should also imagine that these capsids would sometimes contain FUCA-like entities capable to work as ribosomes, but also that most progenotes might lack the translation systems and use sister FUCA-like systems to express their genetic content. In any case, it is possible to imagine some of the first viruses as having the complete set of mechanisms capable to allow their replication, being therefore autonomous. This way, we would not be surprised whether a virus containing the entire translational machinery will be found in a near future. This will provide definitive evidence that viruses can also be autonomous, even if the great amount of them have taken the evolutionary decision of not carrying these enormous energy-consuming nano-machines.

### 13. The age of organisms: the emergence of cells

Along the evolution of progenotes, we envision different populations assembling biochemical pathways independently according to the specific binding properties of peptides. For example, some progenotes with a duplicated genome may have suffered random mutations in their encoding genes that allowed the emergence of a protein containing a new sort of domain capable to bind carbohydrates with more affinity than before. This will probably raise their fitness and may help in a quicker cycling of carbohydrates, making them more readily available in the media. As the nucleotide has a carbohydrate core formed by a ribose, this event can also cascade to a faster production of nucleotides, mainly ATP, that is not only an adenine nucleotide but also the most important energetic coin at the molecular world. On the other hand, other populations of progenotes may evolve into getting a specialized binding of amino acids, while others specialized in binding nucleic acids, and another specialized in binding carbon metabolites that provided carbon for their replication, among other atoms and ions. Those novelties in peptide binding caused by genome duplications, mutations and neofunctionalization processes allowed the parallel evolution of biochemical pathways in the age of progenotes. Importantly, some progenotes' populations evolved proteins capable to bind lipids and these newly formed lipid-binding proteins will end up in producing a new a very successful strategy of life: the cellular one.

When we compare the genomes of organisms from the two great lineages of cells we know today: bacteria and archaea, we observe that most proteins of the basal machinery are highly similar. This level of similarity allows researchers to infer homology, suggesting that bacteria and archaea share a set of genes of common origin, derived from the progenote population that gave rise to these lineages. According to the work of Carl Woese, the ancestral among bacteria, archaea and eukarya has been named LUCA (Woese et al., 1990). We know today that eukaryotes were originated by a new level of symbiotic interaction between an archaea from the Lokiarchaeota clade and a bacteria (Zaremba-Niedzwiedzka et al., 2017). Eukaryotes inherited the genome of archaeas and the membranes of bacteria, together with the endosymbiosis of a bacteria that became the mitochondria.

Thus, the conservation analysis of nucleotides and amino acids between archaea and bacteria evidences their ancestry relationship, but invalidates prokaryotes as a cellular entity without nuclear membrane, in accordance with other works (Di Giulio, 2015, 2023). For the fact that the homology of hundreds of genes encoding proteins from basic cell machinery between bacteria and archaea are known, LUCA is therefore considered to be this progenote ancestor that assembled those shared biochemical paths of homolog enzymes in bacteria and archaea (de Farias et al., 2021). In addition to the sharing of genes between the basal lineages, we can also observe the sharing of protein motifs and domains, suggesting that the formation of these lineages took profit of the recombination of sequences that increased the genetic repertoire (Di Giulio, 2008; Sato and Atomi, 2011). However, for at least two important paths, archaeal and bacterial enzymes are not homologous and their genes and proteins present considerable divergence in their sequence and structure similarity. Thus, even if their functions are analogous, their origin was probably independent. These are the paths for (i) lipid biosynthesis and (ii) DNA replication.

### 14. The age of organisms: the origin of cellular membranes

Inspired by the original work of Woese et al. (1990), many authors suggest a single origin for the two basal domains of life on Earth: Archaea and Bacteria, and the genetic and molecular constitution of their common ancestor (LUCA) has been extensively studied (Bada and Lazcano, 2002; Weiss et al., 2016). Most of these studies consider LUCA as a cellular entity presenting a lipidic membrane and a complex genome formed by double-strand DNA. However, some authors have been questioning the cellularity of the LUCA based on the divergence of

enzymes used for two main pathways: (i) the biosynthesis of membrane's phospholipids; and (ii) the biosynthesis of DNA (Sojo, 2019; de Farias et al., 2021; Prosdocimi and Farias, 2023).

Since the 1990s, knowledge about the constitution of the plasma membrane of Archaea has sparked various discussions on how differentiation between basal cellular groups could have occurred (Lombard and Moreira, 2011). The most important understanding about the constitution of cellular membranes is that they have never been exclusively phospholipid-based (de Farias et al., 2021). Even today, cell membranes contain almost 40% of protein content, with proteins playing crucial roles in the membrane as receptors, transporters, and signal transducers, among others (Farias and Prosdocimi, 2022). The membrane, therefore, would have emerged from the evolution of lipid-binding proteins (Sojo, 2019) that initially were more proteinaceous than lipidic, possibly resembling more of a viral capsid (Prosdocimi and Farias, 2023). As every biochemist know, lipid-binding proteins originated from peptides containing closely related hydrophobic amino acid residues in their 3D structure. These residues function as anchors, allowing proteins stack inside lipid droplets. Along the evolution of various families, lipid-binding proteins containing those sets of hydrophobic amino acids acquired new domains and became capable of binding other molecules in the primordial soup. The appearance of those extra and intracellular domains allowed new signaling macromolecules to emerge (Farias et al., 2021). At the origin of cells, the most significant molecular events were probably the maturation of these macromolecules, allowing the emergence of a controlled interaction between the inside of newly formed cells and the outside environment. Through the evolution of transmembrane protein receptors, the process of macrocode auto-organization constituted the initial steps after the evolution of compartmentalization. Thus, the cells were learning how to communicate the inside to the outside, how outside signals should modify and allow a response from the inside and how to self-organize this new system. This learning was guided by natural molecular selection operating both positive and negative selection, as the most fitted cells survived and the less fitted died. Although progenotes were "naked", free-living RNP macromolecular arrangements, the concept of an individual organism could not apply. However, progenotes also presented some sort of pseudo compartmentalization given by the surrounding set of RNPs that assisted their stability, such as nowadays RNP sets allow the phenomenon of liquid-liquid phase separation inside cells through the formation of aggregates (Prosdocimi and Farias, 2023). Along the emergence of cells, those pseudo compartments will evolve to form a strong barrier between the inside and the outside advancing into the age of organisms. It is reasonable to image that the first membranes were less selective and presented more "open" pores. Over the evolution of cells and with the emergence of new macromolecules, these membranes and their proteins would become increasingly specific, both in binding molecules from outside and in producing organized internal responses to these bindings.

To explain why bacteria and archaea have different constitution of phospholipids, some authors suggested that their cellular ancestor (LUCA) exhibited a mixture of the main phospholipids found in current prokaryotic membranes (Wächtershäuser, 2003). They suggest that this constitution conferred an adaptive value, and that the differentiation happened after the separation basal clades, when selective pressures lead to a purification process among these lineages. However, studies evaluating the properties of such mixed types of membranes have shown that they exhibited similar characteristics to those observed in the modern organisms, refuting the idea of an increased adaptive value (Shimada and Yamagishi, 2011). Also, comparative genomics analyses evidenced that the lipid biosynthesis pathways were completely different and exclusive in archaea and bacteria, with some exceptions observed in specific groups that showed traces of lateral gene transfer events (Coleman et al., 2019). We interpret this exclusivity by the independent origins of archaeal and bacterial membranes from different progenote populations (Prosdocimi and Farias, 2023).

## 15. The age of organisms: the origin of DNA genomes

Many authors suggested that the common ancestor between Archaea and Bacteria was not a cellular entity, but rather a descendant of a population of non-cellularized progenotes (Gogarten and Deamer, 2016; Di Giulio, 2021; De Farias et al., 2021). As we saw, DNA biosynthetic enzymes do not show significant sequence similarity to infer homology, even if there are homologies in pathways involving RNA metabolism. These data suggest that the information was still stored in RNA molecules in progenotes (Prosdocimi and Farias, 2023). Some authors, despite accepting a common origin for cellularization, suggest that the first cellular lineages still possessed an RNA-based genome and that the transition to a DNA-based genome would have occurred independently in these lineages, with viruses playing important roles throughout this process (Forterre, 2005). With the popularity of the RNA world model (Gilbert, 1986), the idea that biological information was initially stored in RNA molecules is now widely accepted. Thus, it is necessary to explain how the transition from RNA to DNA must have occurred, describing which entities and mechanisms were evolved. For accomplishing this goal, the study of polymerases capable of mediating RNA to DNA biosynthesis is essential.

This issue has been tackled by a series of studies by Farias and collaborators. In 2017, these authors identified similarities between (i) concatenated and translated sequences of proto-tRNAs with (‘) RNA-dependent RNA polymerase enzymes (de Farias et al., 2017). This adds further evidence to the tRNA-core hypothesis (de Farias et al., 2016), which suggests that tRNAs contributed to the emergence of the first encoded biological information, being one of the first genes to become fixed during FUCA’s maturation (Prosdocimi et al., 2019). By analyzing the various enzymes with nucleic acid polymerase activity, these authors observed sequence and structural similarity, indicating that they diverged from a common ancestor (de Farias et al., 2017). Thus, it was suggested that RNA-dependent DNA polymerases (reverse transcriptases) were direct descendants of RNA-dependent RNA polymerases. The emergence of these DNA polymerase enzymes allowed the synthesis of the first DNA molecules, which began to spread throughout the progenote’ RNA genomes due to its greater stability. It was found that first DNA polymerases exhibited low processivity and lacked proofreading activity, leading to the insertion of numerous errors during the replication and retrotranscription processes.

In another work, Farias et al. (2023) analyzed the phylogeny of the three families of the main replicative DNA polymerases: families A, B, and C. DNA polymerases from family A can be only found in bacteria, mitochondria, and chloroplasts. In bacteria, these enzymes are involved in resolving Okazaki fragments, while in mitochondria and chloroplast they are involved in whole genome replication. DNA polymerases from family C are also found in bacteria, although they are involved in nuclear genome replication. Both families A and C still today present some residual retrotranscription activities, suggesting an origin from this class of RNA dependent DNA polymerases (Farias et al., 2023).

On the other hand, DNA polymerases from family B constitute the primary replicative DNA polymerases of archaeal and eukaryotic genomes. Although all enzymes from different families perform essentially the same catalytic activity, they do not exhibit sufficient sequence similarity to infer homology, indicating that they may have emerged through multiple and independent evolutionary processes, possibly originating from reverse transcriptases. They also reconstructed the ancestral sequence of DNA polymerases from the B family and showed that these ancestral molecules had even more structural similarities with RNA-dependent DNA polymerases than with modern examples of DNA polymerases B. This result led them to corroborate the idea that DNA polymerases emerged from modifications in reverse transcriptase-like enzymes, further suggesting that the proofreading activity of DNA polymerases may have emerged through modifications in the exonuclease domain of reverse transcriptases (de Farias et al., 2023). This exonuclease domain is important for the degradation of the RNA

molecule after the synthesis of a complementary DNA molecule. Together, these findings suggested that DNA biosynthesis originated independently at least three times in cellular lineages (Farias et al., 2023). Their data indicate that families A and C emerged in distinct groups of bacteria and, after their emergence, were disseminated to other groups through viral lineages. Members of family B probably emerged in a viral group and were subsequently transferred to an archaeal ancestor.

## 16. Concluding remarks

Here, we have reviewed recent literature in the origins of life and proposed a scenario with four ages and three major transitions for describing the evolutionary path from molecules to cells. We traced the main events based on logical assumptions, taking on account gradualism to provide a model with stepwise advances of self-organization forces to gain in the complexification scale. Although self-organization seems to violate the entropy principle and the second law of thermodynamics, the mere existence of living organisms needs explanation. In fact, while the second law of thermodynamics states that the total entropy of an isolated system tends to increase over time, it does not preclude the formation of ordered structures in specific parts of that system, provided there is compensation for an increase in entropy elsewhere. Also, in far-from-equilibrium systems, such as reactive chemical systems or biological systems consuming energy, patterns of self-organization can spontaneously arise due to nonlinear interactions among the components of the system.

In the current model, the first era started with the formation of the basic biomolecules and developed further until these molecules have become abundant in the prebiotic soup (Prosdocimi and Farias, 2023). This is a necessary step to allow the encounter between two chemical worlds previously separated in prebiotic Earth. At some point, nucleic acids and peptides connected for good, starting a successful relationship of replication and encoding, allowing each to endure more, cycle quicker and with better precision. This event was named as chemical symbiosis and the mutualist molecular behavior self-organized the origin of the most important biological code (Prosdocimi et al., 2019, 2021). The emergence of the genetic code, with its precise attributions of codons to amino acids, was a major event that inaugurated earthly life. It entailed the integrated coordination of events and the complexification of molecular interactions to an unprecedented degree in the physico-chemical realm. This event originated the first living organism, and the FUCA has already born as a population of self-replicating RNAs closely linked to amino acids and peptides that became capable to bind amino acids together (Prosdocimi et al., 2019). The spark that ignited life was the emergence of catalytic potential within a ribozyme, the proto-PTC. The proto-PTC would eventually evolve into the complex RNP nano-machine known as the ribosome (Petrov et al., 2014). Along this trajectory on which biology unfolded, it happened the emergence of the most significant early organic code and the development of the protein synthesis apparatus (Farias et al., 2021; Farias and Prosdocimi, 2022).

Matured FUCA populations would contain small RNA genomes with information still mixed with the own ribosomal RNA. But then, genomes will start to duplicate by head-to-tail concatenation and recombine, producing different populations of RNP entities. Some of these subsystems will specialize in produce proteins while others will become free to explore the space of interactions in the primitive soup. However, we must acknowledge that biology cannot exist properly without the access to a FUCA-like translator system, because biology is a science of codes and those codes needs decoder systems. Thus, we envision a rich chemical environment full of RNPs of different populations with lots of translation centers capable to be accessed by non-coding progenotes. Those microenvironments were stabilized by certain types of RNPs that allow a sort of aggregation and allowed a primitive sort of compartmentalization (Zhao and Zhang, 2020). As long as progenotes would reproduce and conquer new microenvironments, new encoded proteins

will be capable to bind other molecules in the soup and, little by little, hardwire protobiotic reactions and pathways (Prosdocimi and Farias, 2022).

The efficiency in binding molecules and catalyze metabolic transformations happened in specific populations of progenotes and were further shared by lateral transference amongst them. The origin of organisms happened when specific sorts of protein allowed the formation of peptidic capsids capable to involve a part of the RNPs, including the translator, and facilitate their transportation to close environments (de Farias et al., 2019; Prosdocimi et al., 2023). This viral-like displacement allowed progenotes to seed life elsewhere around the place of its original unfolding.

At some point, one entity assembled most of biochemical pathways evolved in different progenote populations, giving rise to the prokaryotic ancestor (Farias and Prosdocimi, 2022). LUCA was probably self-organized in a single RNA genome associated with RNPs, having a number of protein-coding genes measured in hundreds, and being capable to bind and metabolize carbohydrates, phosphates, small carbon compounds, cofactors, ions and other metabolites (Weiss et al., 2016). LUCA however was still incapable to interact closely the long chains of carbons that constitute lipid molecules. We can also suppose that the initial environment on which FUCA and the progenotes were born did not had the presence of lipids. This fact would justify both the fact of lipid biosynthesis being one of the last paths to be incorporate and the importance of viral-like strategy to allow this geographical shift.

Nevertheless, when progenotes contacted an environment with abundant lipids, peptides containing hydrophobic amino acids were capable to interact with those molecules and start to hardwire protobiotic transformations. This probably happened in different though close microenvironments, allowing the evolution of two parallel paths for producing bacterial-like membranes and archaeal-like ones (Sojo, 2019). The most important evolution happening in those days would be the signaling paths on which outside molecular information should trigger a cascade of events in the intracellular medium capable to respond to the original stimulus. It is still a matter of debate how differently bacteria and archaea trigger response mechanisms to the binding of ligands.

Finally, studying the history of DNA polymerases, it has been possible to also describe two possible origins for this nucleic acids' biosynthesis (Farias et al., 2023). With these two analogous mechanisms for producing DNA, the emergence of life stabilized in those two strategies (viral and cellular). While the viral architecture probably arose multiple times, cells were originated as fewer members. Besides, bacterial and archaeal basic architectures, some studies suggest the early emergence of other basic clades that were now extinct, like a putative mitochondrial clade. Although we wait more evidence to consider this issue, we know that bacteria and archaeal members will enroll in another symbiotic event, this time happening in an outer ring of the accretion onion, another scale of the fractal. While the origin of (cellular) life is often dated in 3,8 billion years, it then took about 2 billion years later for the next transition to happen. This will show up when some bacteria and archaea (maybe assisted by viruses) will collaborate to create a new sort of cellular strategy, a cell that will be better divided in compartments. The origin eukaryotes came to be either when the cellular membrane invaginate or, alternatively, when multiple events of engulfing happened together controlled by a single central of genetic information, the cellular nucleus. The relevance of eukaryotic constitution is that, after 1,5 billion years of evolution, this organization will allow the origin of multicellularity. But these events happened when evolution was occurring after the cellularization and has been treated extensively elsewhere (Maynard-Smith and Szathmáry, 1995).

## Statement

During the preparation of this work the authors used (i) ChatGPT 3.5 in order to improve the quality and efficiency of the writing process, and (ii) Midjourney bot to provide significant pictures to represent the

concepts described. After using these services, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

## CRediT authorship contribution statement

**Francisco Prosdocimi:** Writing – original draft, Project administration, Funding acquisition, Data curation, Conceptualization. **Sávio Torres de Farias:** Writing – review & editing, Supervision, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We thank Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) for providing research productivity fellowships for FP (CNE E-26/200.940/2022 and 306346/2022-2) and STF (313844/2021-6).

## References

- Almendral, J.M., 2013. Assembly of simple icosahedral viruses. *Subcell. Biochem.* 68, 307–328. [https://doi.org/10.1007/978-94-007-6552-8\\_10](https://doi.org/10.1007/978-94-007-6552-8_10). PMID: 23737056.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215 (3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2). PMID: 2231712.
- Baaske, P., Weinert, F.M., Duhr, S., Lemke, K.H., Russell, M.J., Braun, D., 2007. Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proc. Natl. Acad. Sci. U. S. A.* 104 (22), 9346–9351. <https://doi.org/10.1073/pnas.0609592104>. Epub 2007 May 9. PMID: 17494767; PMCID: PMC1890497.
- Bäckström, D., Yutin, N., Jørgensen, S.L., Dharamshi, J., Homa, F., Zaremba-Niedwiedzka, K., Spang, A., Wolf, Y.I., Koonin, E.V., Ettema, T.J.G., 2019. Virus genomes from deep sea sediments expand the ocean megavirome and support independent origins of viral gigantism. *18 mBio* 10 (2), e02497. <https://doi.org/10.1128/mBio.02497-18>. PMID: 30837339; PMCID: PMC6401483.
- Bada, J.L., 2013. New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. *Chem. Soc. Rev.* 42 (5), 2186–2196. <https://doi.org/10.1039/c3cs35433d>. Epub 2013 Jan 22. PMID: 23340907.
- Bada, J.L., Lazcano, A., 2002. Miller revealed new ways to study the origins of life. *Nature* 416 (6880), 475. <https://doi.org/10.1038/416475a>. PMID: 11932715.
- Barbieri, M., 2003. *The Organic Codes: an Introduction to Semantic Biology*. Cambridge University Press, Cambridge.
- Barbieri, M., 2014. From biosemiotics to code biology. *Biol. Theory* 9, 239–249. <https://doi.org/10.1007/s13752-013-0155-6>.
- Barreat, J.G.N., Katzourakis, A., 2023. A billion years arms-race between viruses, virophages, and eukaryotes. *eLife* 12, RP86617. <https://doi.org/10.7554/eLife.86617>. PMID: 37358563; PMCID: PMC10328495.
- Becker, S., Schneider, C., Okamura, H., Crisp, A., Amatov, T., Dejmek, M., Carell, T., 2018. Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. *Nat. Commun.* 9 (1), 163, 11.
- Bose, T., Fridkin, G., Davidovich, C., Krupkin, M., Dinger, N., Falkovich, A.H., Peleg, Y., Agmon, I., Bashan, A., Yonath, A., 2022. Origin of life: protoribosome forms peptide bonds and links RNA and protein dominated worlds. *Nucleic Acids Res.* 50 (4), 1815–1828. <https://doi.org/10.1093/nar/gkac052>.
- Burton, A.S., Stern, J.C., Elsila, J.E., Glavin, D.P., Dworkin, J.P., 2012. Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. *Chem. Soc. Rev.* 41 (16), 5459–5472. <https://doi.org/10.1039/c2cs35109a>. Epub 2012 Jun 15. PMID: 22706603.
- Carter, C.W. Jr, 2017. Coding of class I and II aminoacyl-tRNA synthetases. *Adv. Exp. Med. Biol.* 966, 103–148. [https://doi.org/10.1007/5584\\_2017\\_93](https://doi.org/10.1007/5584_2017_93). PMID: 28828732; PMCID: PMC5927602.
- Chambers, C.R., Patrick, W.M., 2015. Archaeal nucleic acid ligases and their potential in biotechnology. *Archaea* 2015, 170571. <https://doi.org/10.1155/2015/170571>. PMID: 26494982; PMCID: PMC4606414.
- Coleman, G.A., Pancost, R.D., Williams, T.A., 2019. Investigating the origins of membrane phospholipid biosynthesis genes using outgroup-free rooting. *Genome Biol. Evol.* 11 (3), 883–898. <https://doi.org/10.1093/gbe/evz034>.
- Colson, P., de Lamballerie, X., Fournous, G., Raoult, D., 2012. Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. *Intervirology* 55 (5), 321–332. <https://doi.org/10.1159/000336562>. Epub 2012 Apr 14. PMID: 22508375.

- Colson, P., Levasseur, A., La Scola, B., Sharma, V., Nasir, A., Pontarotti, P., Caetano-Anollés, G., Raoult, D., 2018. Ancestry and mosaicism of giant viruses supporting the definition of the fourth TRUC of microbes. *Front. Microbiol.* 9, 2668. <https://doi.org/10.3389/fmicb.2018.02668>. PMID: 30538677; PMCID: PMC6277510.
- Cornish-Bowden, A., Cárdenas, M.L., 2020. Contrasting theories of life: historical context, current theories. In search of an ideal theory. *Biosystems* 188, 104063. <https://doi.org/10.1016/j.biosystems.2019.104063>.
- Dantas, P.H.L.F., José, M.V., de Farias, S.T., 2021. Structural computational analysis of the natural history of class I aminoacyl-tRNA synthetases suggests their role in establishing the genetic code. *J. Mol. Evol.* 89 (9–10), 611–617. <https://doi.org/10.1007/s00239-021-10029-x>. Epub 2021 Sep 9. PMID: 34505179.
- DasGupta, S., Zhang, S.J., Smela, M.P., Szostak, J.W., 2023. RNA-catalyzed RNA ligation within prebiotically plausible model protocells. *Chemistry* 29 (43), e202301376. <https://doi.org/10.1002/chem.202301376>. Epub 2023 Jun 29. PMID: 37216492.
- Davidovich, C., Belousoff, M., Bashan, A., Yonath, A., 2009. The evolving ribosome: from non-coded peptide bond formation to sophisticated translation machinery. *Res. Microbiol.* 160 (7), 487–492. <https://doi.org/10.1016/j.resmic.2009.07.004>.
- Davidovich, C., Belousoff, M., Wekselman, I., Shapira, T., Krupkin, M., Zimmerman, E., Bashan, A., Yonath, A., 2010. The Proto-Ribosome: an ancient nano-machine for peptide bond formation. *Isr. J. Chem.* 50 (1), 29–35. <https://doi.org/10.1002/ijch.201000012>. PMID: 26207070; PMCID: PMC4508870.
- Davis, B.K., 2002. Molecular evolution before the origin of species. *Prog. Biophys. Mol. Biol.* 79 (1–3), 77–133. [https://doi.org/10.1016/s0079-6107\(02\)00012-3](https://doi.org/10.1016/s0079-6107(02)00012-3). PMID: 12225777.
- de Farias, S.T., 2013. Suggested phylogeny of tRNAs based on the construction of ancestral sequences. *J. Theor. Biol.* 335, 245–248. <https://doi.org/10.1016/j.jtbi.2013.06.033>. Epub 2013 Jul 17. PMID: 23871958.
- de Farias, S.T., Régo, T.G., José, M.V., 2016. tRNA core hypothesis for the transition from the RNA world to the ribonucleoprotein world. *Life* 6 (2), 15. <https://doi.org/10.3390/life6020015>. PMID: 27023615; PMCID: PMC4931452.
- de Farias, S.T., Dos Santos Junior, A.P., Régo, T.G., José, M.V., 2017. Origin and evolution of RNA-dependent RNA polymerases. *Front. Genet.* 8, 125. <https://doi.org/10.3389/fgene.2017.00125>. PMID: 28979293; PMCID: PMC5611760.
- de Farias, S.R., Régo, T.G., José, M.V., 2017. Peptidyl transferase center and the emergence of the translation system. *Life* 7 (2), 21. <https://doi.org/10.3390/life7020021>. PMID: 28441334; PMCID: PMC5492143.
- de Farias, S.T., Jheetka, S., Prosdocimi, F., 2019. Viruses as a survival strategy in the armory of life. *HPLS* 41, 45. <https://doi.org/10.1007/s40656-019-0287-5>.
- de Farias, S.T., Jose, M.V., Prosdocimi, F., 2021. Is it possible that cells have had more than one origin? *Biosystems* 202, 104371. <https://doi.org/10.1016/j.biosystems.2021.104371>. Epub 2021 Jan 30. PMID: 33524470.
- de Farias, S.T., Furtado, A.N.M., Dos Santos Junior, A.P., José, M.V., 2023. Natural history of DNA-dependent DNA polymerases: multiple pathways to the origins of DNA. *Viruses* 15 (3), 749. <https://doi.org/10.3390/v15030749>. PMID: 36992459; PMCID: PMCID: PMC10052633.
- Di Giulio, M., 2008. The origin of genes could be polyphyletic. *Gene* 426 (1–2), 39–46. <https://doi.org/10.1016/j.gene.2008.07.024>. Epub 2008 Jul 29. PMID: 18706983.
- Di Giulio, M., 2015. The non-biological meaning of the term "prokaryote" and its implications. *J. Mol. Evol.* 80 (2), 98–101. <https://doi.org/10.1007/s00239-014-9662-8>. Epub 2014 Dec 23. PMID: 25534641.
- Di Giulio, M., 2021. The phylogenetic distribution of the cell division system would not imply a cellular LUCA but a progenetic LUCA. *Biosystems* 210, 104563. <https://doi.org/10.1016/j.biosystems.2021.104563>. Epub 2021 Oct 13. PMID: 34653531.
- Di Giulio, M., 2023. The absence of the evolutionary state of the Prokaryote would imply a polyphyletic origin of proteins and that LUCA, the ancestor of bacteria and that of archaea were progenotes. *Biosystems* 233, 105014. <https://doi.org/10.1016/j.biosystems.2023.105014>. Epub 2023 Aug 29. PMID: 37652180.
- Dyson, F.J., 1982. A model for the origin of life. *J. Mol. Evol.* 18 (5), 344–350. <https://doi.org/10.1007/BF01733901>.
- Eigen, M., 1971. Self organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465–523. <https://doi.org/10.1007/BF00623322>.
- Eigen, M., Schuster, P., 1977. The hypercycle. A principle of natural self-organization. Part A: emergence of the hypercycle. *Naturwissenschaften* 64 (11), 541–565. <https://doi.org/10.1007/BF00450633>. PMID: 593400.
- Eigen, M., Schuster, P., Sigmund, K., Wolff, R., 1980. Elementary step dynamics of catalytic hypercycles. *Biosystems* 13 (1–2), 1–22. [https://doi.org/10.1016/0303-2647\(80\)90002-7](https://doi.org/10.1016/0303-2647(80)90002-7). PMID: 7437504.
- Erlenbach, N., Grünewald, C., Krstic, B., Heckel, A., Prisner, T.F., 2019. "End-to-end" stacking of small dsRNA. *RNA* 25 (2), 239–246. <https://doi.org/10.1261/rna.068130.118>. Epub 2018 Nov 7. PMID: 30404925; PMCID: PMC6348986.
- Farias, S.T., Prosdocimi, F., 2022. RNP-world: the ultimate essence of life is a ribonucleoprotein process. *Genet. Mol. Biol.* 45 (3 Suppl. 1), e20220127. <https://doi.org/10.1590/1678-4685-GMB-2022-0127>. PMID: 36190700; PMCID: PMC9528728.
- Farias, S.T., Moreira, C.H.C., Guimaraes, R.C., 2007. Structure of the genetic code suggested by the hydrophathy correlation between anticodons and amino acid residues. *Orig. Life Evol. Biosph.* 37, 83–103. <https://doi.org/10.1007/s11084-006-9008-7>.
- Farias, S.T., Régo, T.G., José, M.V., 2014. Origin and evolution of the peptidyl transferase center from proto-tRNAs. *FEBS Open Bio.* 4, 175–178. <https://doi.org/10.1016/j.fob.2014.01.010>. PMID: 24649398; PMCID: PMC3953717.
- Farias, S.T., Prosdocimi, F., Caponi, G., 2021. Organic codes: a unifying concept for life. *Acta Biotheor.* 69 (4), 769–782. <https://doi.org/10.1007/s10441-021-09422-2>.
- Farias, S.T., Furtado, A.N.M., Santos Junior, A.P.D., José, M.V., 2023. Evolutionary origin of B family DNA-dependent DNA polymerases from retrotranscriptases. *Acta Biotheor.* 69 (4), 769–782. <https://doi.org/10.1007/s10441-021-09422-2>.
- Forterre, P., 2005. The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. *Biochimie* 87 (9–10), 793–803. <https://doi.org/10.1016/j.biochi.2005.03.015>.
- Freeland, S.J., Hurst, L.D., 1998. The genetic code is one in a million. *J. Mol. Evol.* 47 (3), 238–248. <https://doi.org/10.1007/pl00006381>. PMID: 9732450.
- Ganti, T., 2003. The Principles of Life. Oxford University Press, Oxford. <https://doi.org/10.1093/acprofoso/9780198507260.001.0001>.
- Gilbert, W., 1986. The RNA world. *Nature* 319, 618.
- Gogarten, J.P., Deamer, D., 2016. Is LUCA a thermophilic progenote? *Nat. Microbiol.* 1, 16229. <https://doi.org/10.1038/nmicrobiol.2016.229>. PMID: 27886195.
- Gomes, E., Shorter, J., 2019. The molecular language of membraneless organelles. *J. Biol. Chem.* 294 (18), 7115–7127. <https://doi.org/10.1074/jbc.TM118.001192>. Epub 2018 Jul 25. PMID: 30045872; PMCID: PMC6509512.
- Guimaraes, R.C., Moreira, C.H., de Farias, S.T., 2008. A self-referential model for the formation of the genetic code. *Theor. Biosci.* 127 (3), 249–270. <https://doi.org/10.1007/s12064-008-0043-y>.
- Helmbrecht, V., Weingart, M., Klein, F., Braun, D., Orsi, W.D., 2023. White and green rust chimneys accumulate RNA in a ferruginous chemical garden. *Geobiology* 21 (6), 758–769. <https://doi.org/10.1111/gbi.12572>. Epub 2023 Aug 24. PMID: 37615250.
- Igamberdiev, AU., Kleczkowski, LA., 2023. Toward understanding the emergence of life: A dual function of the system of nucleotides in the metabolically closed autopoietic organization. *Biosystems* 104837. <https://doi.org/10.1016/j.biosystems.2023.104837>. Epub 2023 Jan 14. PMID: 36649884.
- Janzen, E., Shen, Y., Vázquez-Salazar, A., Liu, Z., Blanco, C., Kenchel, J., Chen, I.A., 2022. Emergent properties as by-products of prebiotic evolution of aminoacylation ribozymes. *Nat. Commun.* 13 (1), 3631. <https://doi.org/10.1038/s41467-022-31387-0>. PMID: 35752631; PMCID: PMC9233669.
- José, M.V., Morgado, E.R., Bobadilla, J.R., 2023a. Groups of symmetries of the two classes of synthetases in the four-dimensional hypercubes of the extended code type II. *Life* 13 (10), 2002. <https://doi.org/10.3390/life13102002>. PMID: 37895385; PMCID: PMC10607949.
- José, M.V., Bobadilla, J.R., Zamudio, G.S., de Farias ST., 2023b. Symmetrical distributions of aminoacyl-tRNA synthetases during the evolution of the genetic code. *Theor. Biosci.* 142 (3), 211–219. <https://doi.org/10.1007/s12064-023-00394-0>. Epub 2023 Jul 5. PMID: 37402895; PMCID: PMC10423125.
- Kauffman, SA., 1986. Autocatalytic sets of proteins. *Journal of Theoretical Biology* 119, 1–24.
- Keller, M.A., Turchyn, A.V., Ralser, M., 2014. Non-enzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible Archean ocean. *Mol. Syst. Biol.* 10 (4), 725. <https://doi.org/10.1002/msb.20145228>. Published 2014 Apr 25.
- Kim, H.J., Kim, J., 2019. A prebiotic synthesis of canonical pyrimidine and purine ribonucleotides. *Astrobiology* 19 (5), 669–674. <https://doi.org/10.1089/ast.2018.1935>.
- Kirschning, A., 2021. Coenzymes and their role in the evolution of life. *Angew. Chem. Int. Ed. Engl.* 60 (12), 6242–6269. <https://doi.org/10.1002/anie.201914786>. Epub 2020 Oct 23. PMID: 31945250; PMCID: PMC7983987.
- Koga, S., Williams, D.S., Perriman, A.W., Mann, S., 2011. Peptide-nucleotide microdroplets as a step towards a membrane-free protocell model. *Nat. Chem.* 3 (9), 720–724. <https://doi.org/10.1038/nchem.1110>. PMID: 21860462.
- Koonin, E.V., Yutin, N., 2010. Origin and evolution of eukaryotic large nucleocytoplasmic DNA viruses. *Intervirology* 53 (5), 284–292. <https://doi.org/10.1159/000312913>. Epub 2010 Jun 15. PMID: 20551680; PMCID: PMC2895762.
- Krishnamurthy, R., Goldman, A.D., Libeck, D.A., Rogers, K.L., Tor, Y., 2022. Nucleobases in meteorites to nucleobases in RNA and DNA? *J. Mol. Evol.* 90 (5), 328–331. <https://doi.org/10.1007/s00239-022-10069-x>. Epub 2022 Aug 12. PMID: 35960316.
- Kuhn, H., Frank-Kamenetskii, M.D., 2005. Template-independent ligation of single-stranded DNA by T4 DNA ligase. *FEBS J.* 272 (23), 5991–6000. <https://doi.org/10.1111/j.1742-4658.2005.04954.x>. PMID: 16302964.
- Lai, Y.C., Liu, Z., Chen, I.A., 2021. Encapsulation of ribozymes inside model protocells leads to faster evolutionary adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2025054118. <https://doi.org/10.1073/pnas.2025054118>.
- Lanier, K.A., Petrov, A.S., Williams, L.D., 2017. The central symbiosis of molecular biology: molecules in mutualism. *J. Mol. Evol.* 85 (1–2), 8–13. <https://doi.org/10.1007/s00239-017-9804-x>.
- Lei, L., Burton, Z.F., 2023. The 3 31 nucleotide minihelix tRNA evolution theorem and the origin of life. *Life* 13 (11), 2224. <https://doi.org/10.3390/life1311224>. PMID: 38004364; PMCID: PMC10672568.
- Lombard, J., Moreira, D., 2011. Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol. Biol. Evol.* 28 (1), 87–99. <https://doi.org/10.1093/molbev/msq177>.
- Luisi, P.L., 2014. Prebiotic metabolic networks? *Mol. Syst. Biol.* 10 (4), 729. <https://doi.org/10.1002/msb.2014531>. PMID: 24771086; PMCID: PMC4023397.
- Malagoda Pathiranage, K., Martin, C.T., 2023. A simple approach to improving RNA synthesis: salt inhibition of RNA rebinding coupled with strengthening promoter binding by a targeted gap in the DNA. *Methods Enzymol.* 691, 209–222. <https://doi.org/10.1016/bs.mie.2023.06.001>. Epub 2023 Jul 7. PMID: 37914447.
- Maturana, H.R., Varela, F., 1973. De Máquinas Y Seres Vivos. Editorial Universitaria, Santiago, Chile.
- Maurette, M., 1998. Carbonaceous micrometeorites and the origin of life. *Orig. Life Evol. Biosph.* 28 (4–6), 385–412. <https://doi.org/10.1023/a:1006589819844>. PMID: 10357645.
- Maynard Smith, J., Szathmáry, E., 1995. The Major Transitions in Evolution. Freeman, Oxford.

- Mietzsch, M., Pénzes, J.J., Agbandje-McKenna, M., 2019. Twenty-five years of structural parvovirology. *Viruses* 11 (4), 362. <https://doi.org/10.3390/v11040362>. PMID: 3100002; PMCID: PMC6521121.
- Miller, S.L., 1953. A production of amino acids under possible primitive earth conditions. *Science* 117 (3046). <https://doi.org/10.1126/science.117.3046.528>, 528–9, PMID: 13056598.
- Mizuchi, R., Furubayashi, T., Ichihashi, N., 2022. Evolutionary transition from a single RNA replicator to a multiple replicator network. *Nat. Commun.* 13 (1), 1460. <https://doi.org/10.1038/s41467-022-29113-x>. PMID: 35304447; PMCID: PMC8933500.
- Muchowska, K.B., Chevallot-Beroux, E., Moran, J., 2019a. Recreating ancient metabolic pathways before enzymes. *Bioorg. Med. Chem.* 27 (12), 2292–2297. <https://doi.org/10.1016/j.bmc.2019.03.012>.
- Muchowska, K.B., Varma, S.J., Moran, J., 2019b. Synthesis and breakdown of universal metabolic precursors promoted by iron [published correction appears in *Nature*. 2021 May;593(7857):E3]. *Nature* 569 (7754), 104–107. <https://doi.org/10.1038/s41586-019-1151-1>.
- Nussinov, M.D., Otroshchenko, V.A., Santoli, S., 1997. The emergence of the non-cellular phase of life on the fine-grained clayish particles of the early Earth's regolith. *Biosystems* 42 (2–3), 111–118. [https://doi.org/10.1016/s0303-2647\(96\)01699-1](https://doi.org/10.1016/s0303-2647(96)01699-1). PMID: 9184757.
- Parker, E.T., Cleaves, H.J., Callahan, M.P., Dworkin, J.P., Glavin, D.P., Lazcano, A., Bada, J.L., 2011a. Enhanced synthesis of alkyl amino acids in Miller's 1958 H2S experiment. *Orig. Life Evol. Biosph.* 41 (6), 569–574. <https://doi.org/10.1007/s11084-011-9253-2>. Epub 2011 Dec 3. PMID: 22139514.
- Parker, E.T., Cleaves, H.J., Callahan, M.P., Dworkin, J.P., Glavin, D.P., Lazcano, A., Bada, J.L., 2011b. Prebiotic synthesis of methionine and other sulfur-containing organic compounds on the primitive Earth: a contemporary reassessment based on an unpublished 1958 Stanley Miller experiment. *Orig. Life Evol. Biosph.* 41 (3), 201–212. <https://doi.org/10.1007/s11084-010-9228-8>. Epub 2010 Nov 10. PMID: 21063908; PMCID: PMC3094541.
- Parker, E.T., Cleaves, H.J., Dworkin, J.P., Glavin, D.P., Callahan, M., Aubrey, A., Lazcano, A., Bada, J.L., 2011c. Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment. *Proc. Natl. Acad. Sci. U. S. A.* 108 (14), 5526–5531. <https://doi.org/10.1073/pnas.1019191108>. Epub 2011 Mar 21. PMID: 21422282; PMCID: PMC3078417.
- Pearce, B.K.D., Pudritz, R.E., Semenov, D.A., Henning, T.K., 2017. Origin of the RNA world: the fate of nucleobases in warm little ponds. *Proc. Natl. Acad. Sci. U. S. A.* 114 (43), 11327–11332. <https://doi.org/10.1073/pnas.1710339114>. Epub 2017 Oct 2. PMID: 28973920; PMCID: PMC5664528.
- Pedreira-Segade, U., Michot, L.J., Daniel, I., 2018. Effects of salinity on the adsorption of nucleotides onto phyllosilicates. *Phys. Chem. Chem. Phys.* 20 (3), 1938–1952. <https://doi.org/10.1039/c7cp07004g>. PMID: 29297910.
- Pérez-Villa, A., Saitia, A.M., Georgelin, T., Lambert, J.F., Guyot, F., Maurel, M.C., Pietrucci, F., 2018. Synthesis of RNA nucleotides in plausible prebiotic conditions from ab initio computer simulations. *J. Phys. Chem. Lett.* 9 (17), 4981–4987. <https://doi.org/10.1021/acs.jpclett.8b02077>. Epub 2018 Aug 20. PMID: 30107125.
- Petrov, A.S., Williams, L.D., 2015. The ancient heart of the ribosomal large subunit: a response to Caetano-Anolles. *J. Mol. Evol.* 80 (3–4), 166–170. <https://doi.org/10.1007/s00239-015-9678-8>. Epub 2015 Apr 16. PMID: 25877522.
- Petrov, A.S., Bernier, C.R., Hsiao, C., Norris, A.M., Kovacs, N.A., Waterbury, C.C., Stepanov, V.G., Harvey, S.C., Fox, G.E., Wartell, R.M., Hud, N.V., Williams, L.D., 2014. Evolution of the ribosome at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* 111 (28), 10251–10256. <https://doi.org/10.1073/pnas.1407205111>. Epub 2014 Jun 30. PMID: 24982194; PMCID: PMC4104869.
- Pownar, M.W., Gerland, B., Sutherland, J.D., 2009. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459 (7244), 239–242. <https://doi.org/10.1038/nature08013>.
- Prosdocimi, F., Farias, S.T., 2019. A emergência dos sistemas biológicos: uma visão molecular sobre a origem da vida, 1 ed. ArteComCiencia, Rio de Janeiro (In português). <https://www.amazon.com.br/dp/B08WTHJPM4>.
- Prosdocimi, F., Farias, S.T., 2020. From FUCA to LUCA: a theoretical analysis on the common descent of gene families. *Acta Sci. Microbiol.* <https://doi.org/10.31080/ASMI.2020.03.0494>.
- Prosdocimi, F., Farias, S.T., 2021. Life and living beings under the perspective of organic macromolecules. *Biosystems* 206, 104445. <https://doi.org/10.1016/j.biosystems.2021.104445>.
- Prosdocimi, F., Farias, S.T., 2022. Entering the labyrinth: a hypothesis about the emergence of metabolism from protobiotic routes. *Biosystems* 220, 104751. <https://doi.org/10.1016/j.biosystems.2022.104751>. Epub 2022 Aug 5. PMID: 35940497.
- Prosdocimi, F., Farias, S.T., 2023. Origin of life: drawing the big picture. *Prog. Biophys. Mol. Biol.* 180–181, 28–36. <https://doi.org/10.1016/j.pbiomolbio.2023.04.005>. Epub ahead of print. PMID: 37080436.
- Prosdocimi, F., José, M., Farias, S., 2019. The first universal common ancestor (FUCA) as the earliest ancestor of LUCA's (last UCA) lineage. Chapter 3. In: Pontarotti, Pierre (Ed.), Book: Evolution, Origin of Life, Concepts and Methods. Springer. [https://doi.org/10.1007/978-3-030-30363-1\\_3](https://doi.org/10.1007/978-3-030-30363-1_3).
- Prosdocimi, F., Zamudio, G.S., Palacios-Pérez, M., Torres de Farias, S., V José, M., 2020. The ancient history of peptidyl transferase center formation as told by conservation and information analyses. *Life* 10 (8), 134. <https://doi.org/10.3390/life10080134>. PMID: 32764248; PMCID: PMC7459865.
- Prosdocimi, F., José, M.V., de Farias, S.T., 2021. The theory of chemical symbiosis: a marginalian view for the emergence of biological systems (origin of life). *Acta Biotheor.* 69, 67–78. <https://doi.org/10.1007/s10441-020-09388-7>.
- Prosdocimi, F., de Farias, S.T., José, M.V., 2022. Prebiotic chemical refugia: multifaceted scenario for the formation of biomolecules in primitive Earth. *Theor. Biosci.* 141 (4), 339–347. <https://doi.org/10.1007/s12064-022-00377-7>. Epub 2022 Aug 30. PMID: 36042123.
- Prosdocimi, F., Cortines, J.R., José, M.V., Farias, S.T., 2023. Decoding viruses: an alternative perspective on their history, origins and role in nature. *Biosystems* 231, 104960. <https://doi.org/10.1016/j.biosystems.2023.104960>. Epub 2023 Jul 16. PMID: 37437771.
- Ralser, M., 2018. An appeal to magic? The discovery of a non-enzymatic metabolism and its role in the origins of life. *Biochem. J.* 475 (16), 2577–2592. <https://doi.org/10.1042/BCJ20160866>. Published 2018 Aug 30.
- Rimola, A., Sodupe, M., Ugliengo, P., 2019. Role of mineral surfaces in prebiotic chemical evolution. In silico quantum mechanical studies. *Life* 9 (1), 10. <https://doi.org/10.3390/life9010010>. PMID: 30658501; PMCID: PMC6463156.
- Rodin, S.N., Ohno, S., 1997. Four primordial modes of tRNA-synthetase recognition, determined by the (G,C) operational code. *Proc. Natl. Acad. Sci. U. S. A.* 94 (10), 5183–5188. <https://doi.org/10.1073/pnas.94.10.5183>. PMID: 9144212; PMCID: PMC24653.
- Root-Bernstein, M., Root-Bernstein, R., 2015. The ribosome as a missing link in the evolution of life. *J. Theor. Biol.* 367, 130–158. <https://doi.org/10.1016/j.jtbi.2014.11.025>.
- Rossmann, M.G., Johnson, J.E., 1989. Icosahedral RNA virus structure. *Annu. Rev. Biochem.* 58, 533–573. <https://doi.org/10.1146/annurev.bi.58.070189.002533>. PMID: 2673017.
- Sato, T., Atomi, H., 2011. Novel metabolic pathways in Archaea. *Curr. Opin. Microbiol.* 14 (3), 307–314. <https://doi.org/10.1016/j.mib.2011.04.014>. Epub 2011 May 24. PMID: 21612976.
- Shimada, H., Yamagishi, A., 2011. Stability of heterochiral hybrid membrane made of bacterial sn-G3P lipids and archaeal sn-G1P lipids. *Biochemistry* 50 (19), 4114–4120. <https://doi.org/10.1021/bi200172d>.
- Shore, J.A., Holland, B.R., Sumner, J.G., Nieselt, K., Wills, P.R., 2020. The ancient operational code is embedded in the amino acid substitution matrix and aaRS phylogenies. *J. Mol. Evol.* 88 (2), 136–150. <https://doi.org/10.1007/s00239-019-09918-z>. Epub 2019 Nov 28. PMID: 31781936.
- Simmel, F.C., 2007. Towards biomedical applications for nucleic acid nanodevices. *Nanomedicine (Lond.)* 2 (6), 817–830. <https://doi.org/10.2217/17435889.2.6.817>. PMID: 18095848.
- Sojo, V., 2019. Why the lipid divide? Membrane proteins as drivers of the split between the lipids of the three domains of life. *Bioessays* 41 (5), e1800251. <https://doi.org/10.1002/bies.201800251>. Epub 2019 Apr 10. PMID: 30970170.
- Stark, M.R., Rader, S.D., 2014. Efficient splinted ligation of synthetic RNA using RNA ligase. *Methods Mol. Biol.* 1126, 137–149. [https://doi.org/10.1007/978-1-62703-980-2\\_10](https://doi.org/10.1007/978-1-62703-980-2_10). PMID: 24549661.
- Stark, M.R., Pleiss, J.A., Deras, M., Scaringe, S.A., Rader, S.D., 2006. An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *RNA* 12 (11), 2014–2019. <https://doi.org/10.1261/rna.93506>. Epub 2006 Sep 18. PMID: 16983143; PMCID: PMC1624903.
- Stephenson, W., Asare-Okai, P.N., Chen, A.A., Keller, S., Santiago, R., Tenenbaum, S.A., Garcia, A.E., Fabris, D., Li, P.T., 2013. The essential role of stacking adenines in a two-base-pair RNA kissing complex. *J. Am. Chem. Soc.* 135 (15), 5602–5611. <https://doi.org/10.1021/ja310820h>. Epub 2013 Apr 8. PMID: 23517345; PMCID: PMC3664431.
- Suttle, C.A., 2013. Viruses: unlocking the greatest biodiversity on Earth. *Genome* 56 (10), 542–544. <https://doi.org/10.1139/gen-2013-0152>.
- Szilágyi, A., Zachar, I., Scheuring, I., Kun, A., Könnyű, B., Czárán, T., 2017. Ecology and evolution in the RNA world dynamics and stability of prebiotic replicator systems. *Life* 7 (4), 48. <https://doi.org/10.3390/life7040048>. PMID: 29186916; PMCID: PMC5745561.
- Takeuchi, N., Hogeweg, P., 2012. Evolutionary dynamics of RNA-like replicator systems: a bioinformatic approach to the origin of life. *Phys. Life Rev.* 9 (3), 219–263. <https://doi.org/10.1016/j.plrev.2012.06.001>. Epub 2012 Jun 13. PMID: 22727399; PMCID: PMC3466355.
- Ter-Ovanesian, L.M.P., Lambert, J.F., Maurel, M.C., 2022. Building the uracil skeleton in primitive ponds at the origins of life: carbamoylation of aspartic acid. *Sci. Rep.* 12 (1), 19178. <https://doi.org/10.1038/s41598-022-21272-7>. PMID: 36357418; PMCID: PMC9649776.
- Viollet, S., Fuchs, R.T., Munafó, D.B., Zhuang, F., Robb, G.B., 2011. T4 RNA ligase 2 truncated active site mutants: improved tools for RNA analysis. *BMC Biotechnol.* 11, 72. <https://doi.org/10.1186/1472-6750-11-72>. PMID: 21722378; PMCID: PMC3149579.
- Vitas, M., Dobovišek, A., 2018. In the beginning was a mutualism – on the origin of translation. *Orig. Life Evol. Biosph.* 48, 223–243.
- Vörös, D., Könnyű, B., Czárán, T., 2021. Catalytic promiscuity in the RNA World may have aided the evolution of prebiotic metabolism. *PLoS Comput. Biol.* 17 (1), e1008634. <https://doi.org/10.1371/journal.pcbi.1008634>. PMID: 33497378; PMCID: PMC7864428.
- Wächtershäuser, G., 2003. From pre-cells to Eukarya—a tale of two lipids. *Mol. Microbiol.* 47 (1), 13–22. <https://doi.org/10.1046/j.1365-2958.2003.03267.x>.
- Weiss, M.C., Sousa, F.L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., Martin, W.F., 2016. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* 1 (9), 16116. <https://doi.org/10.1038/nmicrobiol.2016.116>. PMID: 27562259.
- Woese, C., 1998. The universal ancestor. *Proc. Natl. Acad. Sci. U. S. A.* 95 (12), 6854–6859. <https://doi.org/10.1073/pnas.95.12.6854>.
- Woese, C.R., Kandler, O., Wheelis, M.L., 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* 87 (12), 4576–4579. <https://doi.org/10.1073/pnas.87.12.4576>.

- Xavier, J.C., Hordijk, W., Kauffman, S., Steel, M., Martin, W.F., 1922. Autocatalytic chemical networks at the origin of metabolism, 2020 Mar 11 Proc. Biol. Sci. 287, 20192377. <https://doi.org/10.1098/rspb.2019.2377>. Epub 2020 Mar 11. PMID: 32156207; PMCID: PMC7126077.
- Xu, J., Appel, B., Balke, D., Wichert, C., Müller, S., 2014. RNA aminoacylation mediated by sequential action of two ribozymes and a nonactivated amino acid. *Chembiochem* 15 (8), 1200–1209. <https://doi.org/10.1002/cbic.201300741>. Epub 2014 Apr 24. PMID: 24764272.
- Xu, J., Chmela, V., Green, N.J., Russell, D.A., Janicki, M.J., Góra, R.W., Szabla, R., Bond, A.D., Sutherland, J.D., 2020. Selective prebiotic formation of RNA pyrimidine and DNA purine nucleosides. *Nature* 582 (7810), 60–66. <https://doi.org/10.1038/s41586-020-2330-9>. Epub 2020 Jun 3. PMID: 32494078; PMCID: PMC7116818.
- Xue, H., Tong, K.L., Marck, C., Grosjean, H., Wong, J.T., 2003. Transfer RNA paralogs: evidence for genetic code-amino acid biosynthesis coevolution and an archaeal root of life. *Gene* 310, 59–66. [https://doi.org/10.1016/s0378-1119\(03\)00552-3](https://doi.org/10.1016/s0378-1119(03)00552-3). PMID: 12801633.
- Yonath, A., Yusupov, G., Yusupova, G., 2000. Crystal structure of the ribosome at 5.5 Å resolution. *Science* 289 (5481), 905–920.
- Yu, Q., Ren, K., You, M., 2021. Genetically encoded RNA nanodevices for cellular imaging and regulation. *Nanoscale* 13 (17), 7988–8003. <https://doi.org/10.1039/d0nr08301a>. PMID: 33885099; PMCID: PMC8122502.
- Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., Stott, M.B., Nunoura, T., Banfield, J.F., Schramm, A., Baker, B.J., Spang, A., Ettema, T.J., 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541 (7637), 353–358. <https://doi.org/10.1038/nature21031>. Epub 2017 Jan 11. PMID: 28077874.
- Zhao, Y.G., Zhang, H., 2020. Phase separation in membrane biology: the interplay between membrane-bound organelles and membraneless condensates. *Dev. Cell* 55 (1), 30–44. <https://doi.org/10.1016/j.devcel.2020.06.033>. Epub 2020 Jul 28. PMID: 32726575.
- Zhuang, F., Fuchs, R.T., Sun, Z., Zheng, Y., Robb, G.B., 2012. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res.* 40 (7), e54 <https://doi.org/10.1093/nar/gkr1263>. Epub 2012 Jan 12. PMID: 22241775; PMCID: PMC3326334.