 doi:10.20944/preprints202003.0419.v5

The ATP hypothesis: a new conceptual framework for the origin of the genetic code

Ping Xie[1, 2]

[1] Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, P.R. China

[2] Yunnan University, Kunming 650500, P. R. China

* Correspondence: xieping@ihb.ac.cn

ABSTRACT

A plenty of theories on the origin of genetic codes have been proposed so far, yet all ignored its energetic relation to the biochemical system. Here, a new hypothesis is proposed, according to which ATP is at the origin of the primordial genetic code by its coevolution with the pristine biochemical system. This hypothesis aims to show how the genetic code was produced by photochemical reactions in a protocell that derived from a lipid vescicle enclosing various life's building blocks (e.g. nucleotides and peptides) . At extant cell, ATP is the only energetic product of photosynthesis, and is at the energetic heart of the biochemical systems. In the prelife vescicle, ATP energetically elongated chains of both polynucleotides and polypeptides, thus providing a bridge between these molecules and eventually mediating biochemical innovation in the protocell from energy transformation to informatization. ATP was not the only one that could drive the formation of polynucleotides and polypeptides, but favored by precellular selection. The protocell innovated the photosynthesis system to produce ATP, which later has been processed efficiently and regularly with the aids of proteins and RNA/DNA. The completion of the genetic code from RNA to DNA marked the dawn of cellular life operated by Darwinian evolution. The ATP hypothesis supports the photochemical origin of life, shedding light on the formation of both photosynthetic and biochemical systems, which remains largely unknown thus far.

Key words: the ATP hypothesis, the origin of the genetic code, life's building block, probiotic "soup", coevolution, biochemical system, energy transformation, informatization, structuralization, precellular selection, photochemical origin of life

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2020 doi:10.20944/preprints202003.0419.v5

INTRODUCTION

The origin of the genetic code is the key to revealing the origin of life on Earth, as it is a prerequisite for the existence of life. The genetic information is recorded in chemical memories (RNA or DNA) by means of triplets of nucleotides named codons, but its origin has been investigated and debated for a long time (e.g., [**1**,**2**,**3**,**4**,**5**]). It is a miracle of nature that a set of genetic codes was able to produce tens of millions of different species on Earth. More than half a century has passed since the discovery of the genetic code, but its origin is still considered to be one of the greatest mysteries in life science. Is the origin of the genetic code truly unknowable? Does the code truly require external design? In this paper, evolution of the genetic code is discussed within the context of the evolution of biochemical reactions in the protocell where ATP was at the energetic center and became the first key building block that initiated a series of potential chain reactions that resulted in triplet codons.

Agnosticism and the deficiencies of the prevalent hypotheses

Most biologists hold a pessimistic view that no one knows exactly how the genetic code came into being, as an exact reconstruction of the process of the construction of genetic code may never be possible [**6**]. Yockey [**7**] claims that the origin of the genetic code is unknowable, as there is no evidence in physics or chemistry of the control of chemical reactions by a sequence of any sort or of a code between sequences. He critically comments that many papers have been published with titles indicating that their subject is the origin of the genetic code, while the content actually deals only with the evolution of the genetic code.

To date, there are several prevalent hypotheses. The frozen accident hypothesis states that the allocation of codons to amino acids in a single ancestor occurred entirely by “chance” and then remained unchanged [**1**]. The stereochemical hypothesis claims that there is, in many cases, a specific stereochemical fit between an amino acid and the base sequence of the corresponding codon on the appropriate tRNA [**8**, **9**]. The stereochemical and frozen accident hypotheses are opposing views. The biosynthetic hypothesis postulates that the code was assigned in parallel to the evolution of amino acid biosynthesis [**2**]. Knight et al. [**10**] declared that the genetic code is a product of selection, history and chemistry. However, none of these describes how genetic codes were originally created within the pristine biochemical system. Since then, very little definitive progress has been made, although the literature is replete with attempts to explain the variation or flexibility of the codes and possible rules of codon allocation to amino acids [**11**, **12**, **13**, **14**].

Frankly speaking, these hypotheses suffer from two defects: first, none of these hypotheses can explain satisfactorily why the genetic code evolved (e.g., energetically) in such a way, and second, none of them can explain the origin of the genetic code in the context of the biochemical system (a relation of part to whole). In other words, all these hypotheses completely overlooked the coevolution of the genetic code with the biochemical system. In my opinion, it is impossible to understand the origin of codons based on the codons alone [**11**, **15**] or even by extending the perspective to the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2020 doi:10.20944/preprints202003.0419.v5

possible relationship between codons and amino acids [13].

In vitro evolution and the chicken-and-egg paradox

Evolution of a primordial cell could have started from a confined environment (e.g. lipid vesicle). Actually, we have long been troubled by a so-called quasispecies model proposed by Eigen [16]. He strongly advocated for the *in vitro* evolution of macromolecules. A quasispecies in the environment was imagined to be a population of genetically related RNA molecules that had certain morphological similarities but were not identical. He hypothesized that the quasispecies followed the Darwinian process of natural selection. This model has been highly influential [6], which, in my opinion, has perhaps misled our understanding of the origin of the genetic code. For example, it is erroneously assumed that there was a random succession of dominant polymer sequences and their associated functional properties in the natural history of life on the primitive Earth [17]. It should be borne in mind that all extant cells are surrounded by a membrane composed of amphipathic lipids, which could favor the biochemical production of key cellular components like ATP. Also, amphipathic lipids can self-assemble in an aqueous medium, providing chemical basis for division of primordial cells.

Molecular biologists have long been troubled by the so-called chicken-and-egg paradox of proteins and nucleic acids, seemingly a logical circular debate about which appeared first ("metabolism first" vs "replication first"). It is also certain that the clues about the origin of macromolecules have been completely lost due to the cyclizing of biochemical pathways in which the transitional states or tracks had long disappeared. Just as the ancient Greek philosopher Heraclitus said, the beginning and ending points overlap on the circumference of a circle. It is great to innovate protein synthesis manipulated by RNA, and vice versa. It is also impossible that such complex coherent pathways and intertwined processes could have evolved separately.

THE ATP HYPOTHESIS

In this article, a new hypothesis—the ATP hypothesis—is presented to explain how ATP played a central role in the origin of genetic code that coevolved with the primordial biochemical system in the protocell (**Figure 1**). ATP was a building block of life in the prebiotic "soup", but currently is renewed by solar-photon driven chemical reactions. In the evolution of the primordial genetic code, ATP served as both an energy carrier and an informatization molecule, as ATP could not only energetically connect amino acids to short peptides and catalytic polypeptides, but also energetically drive the formation of nucleotides, which resulted in thermodynamically favorable and/or probabilistic selection of triplet codons in nucleic acids—RNA and DNA. Eventually, RNA serves as a transient carrier of information, while DNA acts as the final and permanent storage molecule of genetic information by housing the genetic code. The reason why ATP is considered here as a carrier of information is because ATP provides energy for the synthesis of all other nucleotides from simple molecules, while A, G, C, and T/U comprise the genetic code.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2020 doi:10.20944/preprints202003.0419.v5

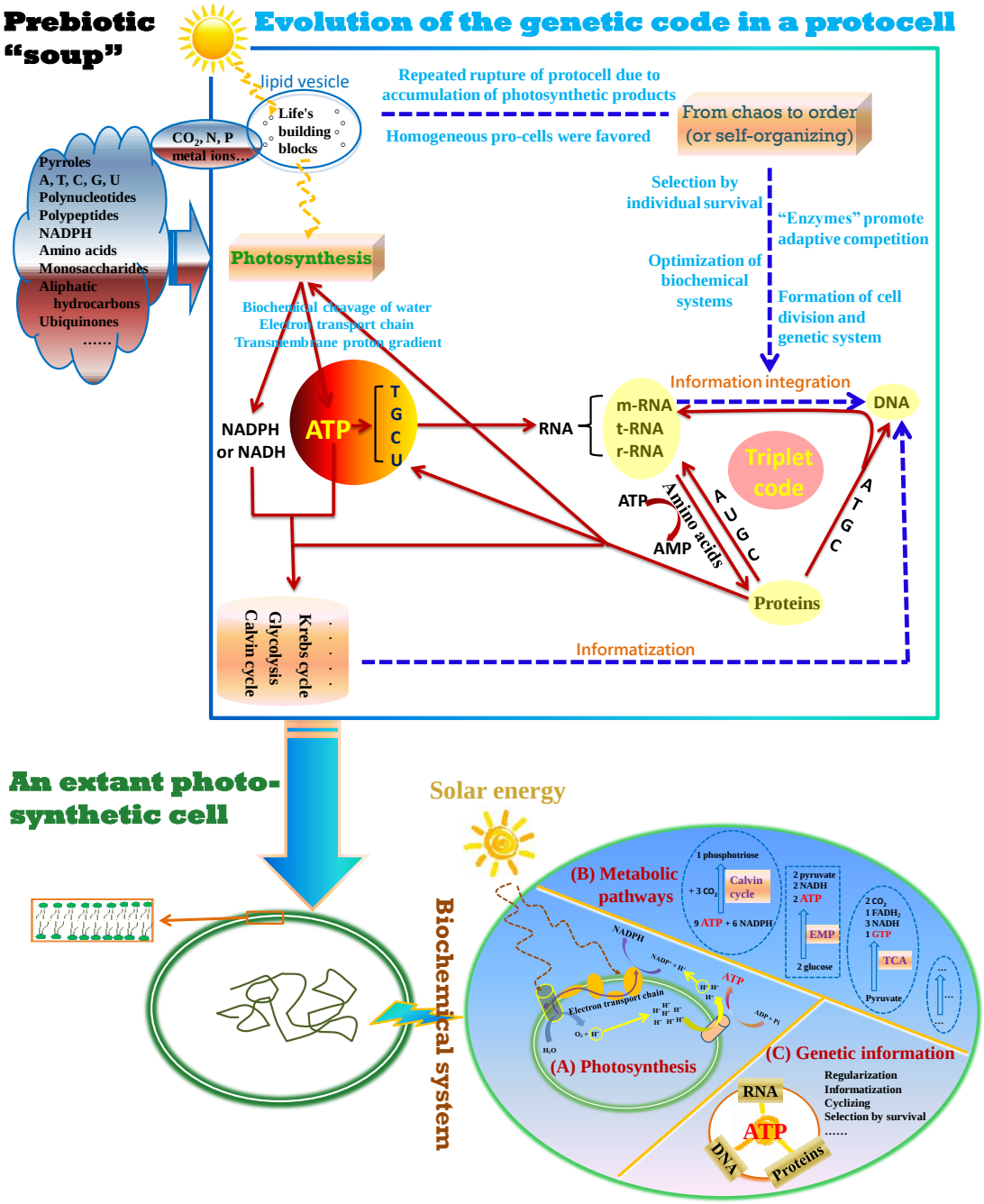

Figure 1. A simplified conceptual model of the origin of the genetic code based on the ATP hypothesis. In the protocell, dashed blue lines indicate evolutionary processes during the precellular period, while solid red lines denote processes or interactions from the precellular period to the present; and arrows indicate the direction of influences or actions. Building blocks of life, impermeable lipid vesicles and energy (e.g. sunlight) are three indispensable elements for the origin of life. The story started from lipid vesicle in the prebiotic "soup" where there were many building blocks of life (e.g., pyrroles, A, T, C, G, U, NADPH, amino acids, aliphatic hydrocarbons, ubiquinones, monosaccharides, etc.) coexisting in sufficient amounts. ATP built a bridge between polynucleotides and polypeptides, leading to precellular

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2020 doi:10.20944/preprints202003.0419.v5

informatization and structuralization. Photochemical production of ATP (biochemical cleavage of water, electron transport chain, transmembrane proton gradient etc.) was selected to meet the origin of the genetic code, a process lasting several hundred million years. Then ATP became the energetic heart through innovating biochemical paths such as Krebs cycle, glycolysis, Calvin cycle etc. It is only the evolutionary completions of cyclizing polynucleotides and polypeptides into a feedback loop of reciprocal causation and the genetic code from RNA to DNA that, contrary to the central dogma, marked the dawn of cellular life, when Darwinian evolution began to operate. The first living cells enclosed in lipid bilayer cell membranes were divisible and replicable, with a delicate biochemical system where solar energy is continuously transformed into ATP through photosynthesis and metabolic pathways, commanded by genetic information in triplet codon.

The key argument favoring ATP as the initiator of the genetic code is its ability to elongate chains of both polynucleotides and polypeptides without additional energy input, making it possible to establish or fix chemical relations between sequences of nucleotides in polynucleotides and amino acids in polypeptides from their numerous random combinations through selection of cellular survival, an ecological force or a feedback mechanism. This is a crucially important point. Otherwise, we cannot explain logically why information could be generated. Technically, photosynthesis, a goal-oriented process, enables various biotic factors or reactions (ATP, lipid vesicles, informatization, structuralization, homogenous individuals, individuality, survival, etc.) to be integrated into an operating system of the genetic code within a poor primordial biochemistry at the beginning. The genetic code is completed with a triplet codon, perhaps only for stereochemical handling of amino acids through, e.g., Watson–Crick pairing interactions, and cyclizing of polynucleotides and polypeptides into a feedback loop of reciprocal causation. Characteristically, the ATP hypothesis can also be viewed as an energetical model for the origin of the genetic code.

How life, with its genetic code, came into being on the primitive Earth more than three billion years ago is an undeniably important but largely conjectural issue. Nevertheless, a logical conjecture can be made in light of certain philosophical perspectives. The ancient Greek philosopher Aristotle once said, ‘we do not have knowledge of a thing until we have grasped its why, that is to say, its cause’. Thus, if, in terms of the Aristotelian philosophy, the causes are as follows: “matter” is composed of various building blocks (e.g., nucleotides, amino acids, sugars, lipids, etc.), “form” makes matter into a particular type of thing (i.e., protocells with biochemical systems including the genetic code, biochemical cycles, etc), the “agents” are ATP produced by photochemical reactions to cause change, and “purpose” is the goal of the form (i.e., individuality of the cell and its desire to struggle for survival).

WHY ATP?

“Hidden hand” or natural forces?

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2020 doi:10.20944/preprints202003.0419.v5

We cannot deny that codons and amino acids are linked stereochemically [9]. Otherwise, we would fall into the religious belief of the genetic code being God's creation or design. The key is that the birth of the genetic code must have been driven by some forces. Randomness and selection were frequently considered to be these forces, but I do believe that they are not the main driving forces. I contend that driving forces should be energetic, e.g., ATP produced by solar-photon driven synthesis of chemicals in the primordial cell.

It is beyond all doubt that the design of genetic codons, a chemical correspondence rule, never required the intervention of a “hidden hand”. However, currently, no reliable fossil evidence is available, and eons of evolution have blurred the molecular vestiges of the early events that remain in living organisms [18]. Fortunately, we can still look back at history from the extant, even if our eyes can perceive only a very minute fraction of the history of early life.

To avoid the “hidden hand”, the initial step for the emergence of the genetic code in a primordial cell should be structurally simple and energetically constant. Subsequently, peptide/nucleic acid interactions and thermodynamic processes became more and more complex, and well-organized with optimized structure, paths and functions under selective pressure.

Inspiration from the organization of the extant biochemical system

The present is a continuation of the past, so to solve the puzzle of the genetic code, it will be fruitful to understand how a biochemical system is organized. Consider the following question: what is life? Life, compositionally, is the unity of matter, energy and information and, dynamically, is the harmonious interplay of material cycling, energy flow and information communication. Energy is the key to supporting life systems. In recent decades, physicists and chemists have discovered many details about how organisms are structured and how they work. It is known that organisms are physiological machines in which inconceivably numerous biochemical reactions take place for the acquisition, conversion and use of energy. Except for chemoautotrophic bacteria, all modern autotrophic life forms (plants, algae and some bacteria) have an energy source in sunlight, which is converted to chemical energy by a series of complex physicochemical processes called photosynthesis [19].

Energetically and informatically, nothing in any biochemical system is more important than ATP. ATP is the only energetic product of photosynthesis, carrying chemical energy converted from sunlight. It then provides energy for metabolism through conversion of ATP/ADP/AMP, supporting the transformation of various biomolecules into each other in an exquisitely organized cell. In other words, the major metabolic pathways (e.g., the Calvin cycle, glycolysis, and Krebs cycle) are all coupled with ATP. Of course, NAD(P)H, a derivative of nucleotides, is also necessary, as it transports H and e^- (through conversion of NAD(P)H/$NAD(P)^+$).

In extant life, in addition to ATP being an indispensable building block of the genetic system (DNA and RNA), the other four nucleotides for genetic coding are all derived from ATP with some of the building blocks of life, as ATP, the only energetic product of photosynthesis, provides biochemical force to these transformations [20].

In this sense, ATP is the only universal currency of biological energy. Information delineates the border between the living and the inanimate, as the living world appears to be the only place where information is recorded, processed, or used [21]. Therefore, as an irreplaceable molecule of both energy transformation and informatization, ATP appears to play energetically a central role in the biochemical system.

Photochemical origin of life

The importance of ATP in biochemical systems could be attributed to its role in governing the evolution of early photosynthetic systems in primordial life, as like all irreversible processes, life must have arisen to dissipate a generalized thermodynamic potential, mostly the solar-photon potential [22]. Although debated, there are signs that life on Earth did start out with photochemical synthesis, as sunlight, needless to say, has been the most universal and constant source of energy, which is particularly important for a process (such as the origination of life) that requires hundreds of millions of years [23].

It is well known that chlorophyll is a very important molecule for photosynthesis. Interestingly, it seems to be an adduct between a magnesium porphyrin ring and a long-chain fatty acid from the “membrane” of a protocell. Both the porphyrin ring and fatty acid were building blocks of life in the “prebiotic soup”. On the other hand, cytochrome is an electron transport protein with iron porphyrin or heme as a prosthetic group and thus is an important part of the primitive photosynthetic system necessary for the production of ATP. The heme was likely derived from a photosynthetic pigment, chlorophyll, as the biosynthetic pathways of these two molecules are very similar in extant life forms [24], although you may argue that the opposite is true. In my opinion, cytochrome was imprinted with photosynthesis. Currently, cytochrome is a universal electronic carrier, present even in chemoautotrophic bacteria [25]. Consider this question: if photosynthetic bacteria were not the last universal common ancestor (LUCA) of all modern life forms, why do chemoautotrophic bacteria use a photosynthesis-imprinted molecule such as cytochrome as an electron carrier?

The thermodynamic dissipation theory for the origin of life assumes that life began and persists today as a catalyst for the absorption and dissipation of sunlight on the surface of Archean seas [26]. This is then strengthened by the factual inference that many fundamental molecules of life are pigments that arose and evolved to dissipate the solar spectrum [22]. The ATP hypothesis further details how life began with the evolution of the biochemical system driven by photochemical synthesis, in which ATP played a key role in the biochemical transformation from energy to information, leading to the birth of the genetic code.

To date, there have been many hypotheses about the origin of life, e.g., the "soup" hypothesis, clay hypothesis, lipid world hypothesis, polyphosphate hypothesis, coenzyme world hypothesis, iron–sulfur world hypothesis, Zn-world hypothesis, PAH world hypothesis, "metabolism first" models, autocatalysis hypothesis, self-organization and replication hypothesis, "deep-hot biosphere" model, deep-sea vent hypothesis, thermosynthesis world hypothesis, radioactive beach hypothesis,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2020 doi:10.20944/preprints202003.0419.v5

ultraviolet and temperature-assisted replication model, bubbles hypothesis, and multiple genesis (see review in [23]). Unlike these hypotheses, the present paper asserts that life arose via photochemical synthesis, i.e., a primitive kind of photosynthesis drove the origin of life, although the most popular current theory on the origin of life is that involving the chemistry of alkaline thermal vents in the sea floor. I admit that there is not yet a general consensus in the origins-of-life community about whether the earliest organisms were photosynthetic, chemosynthetic or heterotrophic. Anyway, the ATP hypothesis is also a new theory about the origin of life, strongly supporting the photochemical origin of life.

It may be argued that prebiotic origin of pyrimidine nucleotides does not depend on ATP, and that these earliest components of nucleotides (even their polymers) can be synthesized chemically, which seems a lot earlier than the appearance of the primitive photosynthesis system [27, 28]. However, the fact that nucleotides were present in the prebiotic soup without the aid of photosynthesis does not deny the possibility that production of ATP coevolved with photosynthesis in the protocell, leading to the completion of the biochemical system as well as its genetic codes in current cells. A, G, C, and T/U or other nucleotides might have occurred in the prebiotic soup, but only ATP was successfully selected as the energetic product of photosynthesis in the protocell. This may either be random or inevitable due to e.g. molecular compatibility. Eventually, with the development of biochemical system, ATP began to drive energetically the formations of other nucleotides using simple molecules such as amino acids, glutamine, one carbon unit, CO_2 etc.

It may also be argued whether this primitive photosynthesis system operate to produce ATP without the involvement of proteins at all and how the photosynthesis system was propagated in early life without involving DNA, or even RNA. Actually, various life' building brocks (e.g. amino acid, purines, pyrimidines) are found to be present in meteorolites [29], indicating that nucleotides could be present or easily formed in the prebiotic soup. The protocell innovated the photosynthesis system to produce ATP, which later has been processed efficiently and regularly with the aids of proteins and RNA/DNA.

HOW ATP?

ATP-mediated biochemical innovation from energy transformation to informatization in the protocell

The genetic code system was built by a chemical mechanism closely related to the production of ATP in the protocell. Energetically, the first task undertaken by primordial life was to achieve sufficient production of this nucleotide. In present-day photoautotrophs, synthesis of ATP requires a transmembrane gradient of protons derived from biochemical cleavage of water in the photosynthetic system. Therefore, to guarantee such a proton gradient, there was first a relatively closed entity or a small space impermeable to H^+. This space was most likely a lipid vesicle, the precursor of the protocell. The fact that synthesis of ATP requires a transmembrane gradient of H^+ made it impossible for macromolecules to evolve *in vitro*, as suggested by Eigen. It was likely an important step for the evolution of life to sequestrate various life's

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2020 doi:10.20944/preprints202003.0419.v5

building blocks in a lipid vesicle where products (e.g. polynucleotides, polypeptides) could accumulate to an immensely higher concentration, which were not free to diffuse away. Also, an enclosed environment helped to protect biochemical reactions of labile compounds.

Chemically, it was not impossible, on the primitive Earth, that fatty acids could automatically form a double-layered globular membrane structure [**30**]. In modern life forms, the cell membrane, consisting of a lipid (phospholipid) bilayer, provides controlled entry and exit ports for the exchange of matter. It permits the passage of small molecules such as CO_2 and O_2 by diffusion but acts as a barrier for certain molecules and ions (e.g., H^+), leading to different concentrations on the two sides of the membrane. H^+ cannot pass freely across the membrane except through transmembrane protein channels. This implies that the formation of ATP in primordial cells gradually began to be dependent on polypeptide channels that then developed into ATP synthase. In addition, the formation of H^+ from the cleavage of H_2O also required the help of polypeptides.

It is reasonable to believe that fluent production of ATP was possible only if the various elements (sunlight, lipid bilayer membrane, polypeptides, cleavage of H_2O, transmembrane H^+ gradient, electron carrier, etc.) had been organized in an orderly manner. It may be inferred that there were a series of events that randomly occurred in protocells. For example, the transmembrane proton gradient coupled with the polypeptide channel resulted in the formation of an apparatus (i.e., ATP synthase) to synthetize ATP. Nucleotides such as ATP could form polynucleotides by self-condensation, some of which carried amino acids (precursor of tRNA), and others built a platform for the synthesis of polypeptides (precursor of rRNA), which eventually replaced the irregular stochastic formation of polypeptides from amino acids activated by ATP. Today, RNA (mRNA, tRNA, rRNA) can carry out key reactions of protein synthesis. On the other hand, the polypeptides in turn participated not only in the construction of the transmembrane channels of hydrophilic molecules/ions but also in the biochemical cleavage of H_2O and in catalyzing the self-condensation of nucleotides. As a consequence, numerous consecutive reactions were linked into a variety of chains, which might be linear, branched, or cyclic. Fundamentally, these were creative/innovative processes representing the creation of order out of disorder and of rationality out of randomness. In short, by these processes, the randomly existing building blocks were finally organized into ordered cellular structures and functions.

Primitive life commenced its journey from materials that were not organized. The prebiotic mixtures of chemicals (also called “prebiotic soup”) present on the primeval Earth were assumed to contain an immense number of building blocks of life, such as amino acids, pyrroles, phosphates, metal ions, nucleotides (e.g., A,T,G,C,U), NADPH, polymeric amino acids, polymeric nucleotides, aliphatic hydrocarbons, plastoquinone, ubiquinones, and monosaccharides [**23**]. It is assumed that these building blocks could form polymers of random sequences or supramolecular aggregates with different properties or functions, such as self-replication and catalysis [**17**].

Let us first consider the impermeable lipid vesicle (enclosed by amphiphiles having

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2020 doi:10.20944/preprints202003.0419.v5

both polar and nonpolar domains) that enclosed sufficient building blocks of life. Their exposure to sunlight would drive the dazzling flow of electrons and H^+, causing active recombination of elements. This included various organic chemical reactions, leading to a series of prebiotic organic syntheses. This would increase the accumulation of large molecules, but accompanied with ceaseless input of small molecules such as CO_2; therefore, the protocells would have had to reciprocate between enlargement and rupture, giving rise to the development of both photochemical synthesis (a precursor of photosynthesis referred to here as quasiphotosynthesis) and cell division. An era of information then followed: in the cycle of quasiphotosynthetic growth followed by division and through the selection of individual survival, the protocell established regularity, reproducibility and rhythmicity of various organic chemical reactions in dealing with photochemically synthetized products, consequently leading to a complex web of biochemical cycles and photosynthesis. This is a process that established information (called informatization) along with cellular compartmentalization (**Figure 1**). Thus, such "reverse micelles" enclosing the building blocks of life were likely the site where life, with its genetic codes, began to emerge, driven by photochemical reactions.

The route from ATP to triplet codon

You may guess that amino acids and nucleotides form giant clusters and that perhaps this is where ATP is involved, i.e., perhaps it can be absorbed into amino acid and/or nucleotide clusters. Random formation of such clusters cannot be excluded, but this was not favored by prebiotic selection. A scenario of how a set of genetic codes was successfully selected to record, preserve and transmit information is outlined below.

① *ATP is a key bridge between polynucleotides and polypeptides*

First, it was likely that in the organic "soup" enclosed by the protocell, the energetic ATP with its derivatives could randomly extend chains of both polynucleotides and polypeptides (as ATP provides biochemical energy for aminoacylation of tRNA by aminoacyl-tRNA synthetases, which is a key prerequisite for the connection of amino acids to form short polypeptides), which made it possible to establish or fix the chemical relation between sequences of nucleotides in polynucleotides and amino acids in polypeptides from their numerous random combinations through selection of cellular survival, an ecological force or a feedback mechanism. Briefly, ATP could do, energetically, two things—extending both nucleic acids (DNA/RNA) and proteins and then making it possible to create a relationship between the two (i.e., a sequence-specific interaction), that is, information (recorded by specific polypeptide-polynucleotide pair)! In this way, ATP mediates biochemical innovation from energy transformation to informatization. This is an indispensable step in the birth of the genetic code and life on Earth and is also the key to the start of the building of biochemical systems, including genetic systems, in the protocell. It should be noted, however, that ATP is not the only one that could drive the formation of polynucleotides and polypeptides, because these polymers can be also synthesized in labs, and thus might exist in prebiotic time. However, this does not negate the role

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2020 doi:10.20944/preprints202003.0419.v5

of ATP as a bridge between polynucleotides and polypeptides, as ATP might much improve these processes. Such ATP-centered genetic/biochemical system was improved and preserved by precellular selection.

② *From informatization to structuralization*

Second, informatization was inevitably coupled with structuralization processes, such as structural subdivision or specialization and functional differentiation, providing a basis for the establishment of the triplet codon system. This is a stage for the development of versatile structures (functional biomolecules). For example, tRNA was specialized to carry specific amino acids, polypeptides helped match the acceptor stem of tRNA to its anticodon, and the system developed the rule of codon-anticodon base pairing, i.e. molecular recognition through stereochemical interactions, e.g., hydrogen bonds, van der Waals forces and aromatic stacking, and ushered in a unified platform, rRNA, for protein synthesis (synthetizing polypeptides according to an mRNA template). In this way, macromolecules became functionally differentiated, i.e., handling (record, preserve and transmit) information via polynucleotides and catalyzing all chemical reactions via polypeptides called enzymes, and both were further cyclized into a system of reciprocal causation. As a result, code-based informatization led to evolutionary innovation of diverse principles or patterns, e.g., biochemical pathways (**Figure 1**). These biochemical processes substantially increased the complexity of the protocell. The foundation of a heritable bio-information system marked a real shift of the world from prebiotic chemistry to primitive biology, and only after this stage did Darwinian evolution begin to operate. Progressive refinement of these mechanisms then provided further selective advantages for protocells.

③ *Why triplets?*

Third, it is reasonable to postulate that the triplet codon was just a result of the optimization of biochemical networks under selective pressure, i.e., for handling more than 20 amino acids, it was not good for there to be either too many (more cumbersome) or too few (4^3 =64, thus there is still considerable redundancy of encoding) codons. That is, three was the lowest number of bases required to encode an amino acid.

④ *Why was heritable homogeneity favored?*

Of course, it must have taken a very long time for protocells to test and modify the genetic code system through the selection of positive phenotypes for precellular survival. Then, the cells obtained the capacity to transmit their blueprint recorded in DNA from generation to generation (replicability), and meanwhile, self-building became a central characteristic of life. The protocell had therefore taken a historic step toward the first genetic cell, i.e., a true species that could bear, process and transmit information, reproducing homogeneous, although not absolutely homogenous, individuals that were thus capable of undergoing Darwinian evolution. It seemed to be a first principle that reproduction of individuals with heritable homogeneity was favored by nature, which—somewhat as a centripetal eco-physiological force—not only governed the integration of various biochemical events but also shaped the direction of survival selection.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2020 doi:10.20944/preprints202003.0419.v5

It is suggested that only macromolecules that operate on their environment to produce further copies of themselves have an evolutionary future [31]. In terms of individuals (protocells), this seems to be a plausible view.

⑤ *From RNA to DNA*

How nucleic acids diverge into DNA and RNA is a great mystery. The primary advantage of DNA over RNA as a genetic material is assumed to be the greater chemical stability of DNA, allowing the formation of much larger genomes based on DNA than on RNA [17], although there are only minor structural differences between the two.

As ATP is the only nucleotide produced by photosynthesis, it was likely not difficult for the cells to convert ATP to any other nucleotides. Meanwhile, these energetic nucleotides could also self-condense into a wide variety of mRNAs, including other RNAs. If the cells could do this, why was it impossible for them to extend the chain to make a DNA molecule? The progress from mRNA to DNA was undoubtedly a great step of informatization in the biochemical system. That is, DNA was specialized to accurately record and permanently preserve all genetic information that provides commands for all cellular activity, while mRNA became short-lived messengers for the implementation of the instructions. It is an astonishing imprint of the evolutionary route from RNA to DNA that RNA is still involved as a primer in the DNA replication of extant organisms.

In the process of information integration, there were subtle structural differences between DNA and RNA. First, the second carbon atom of the ribose was connected to -H in the former but -OH in the latter. Second, thymine (T) in DNA was replaced by uracil (U) in RNA; of course, the structural difference was very small, i.e., T had more than one methyl group. No one knows why only one base was different rather than all four. This is perhaps just because of a need for difference, as similar cases were not rare, e.g., NADPH and NADH. In contrast to DNA, RNA can fold into a variety of complex tertiary structures, analogous to structured proteins, and catalyze a broad range of chemical transformations [32]. It is unclear whether the diverse three-dimensional structures of RNA were due only to such a very small change in the ribose or base. Structurally, was this just an accident? Functionally, subdivision of the genetic system into RNA and DNA might have favored orderly management and control of information in the very small cell in which hundreds of biochemical reactions occurred simultaneously. Consequently, the biochemical system achieved a status where mRNA was immediately destroyed after the task was completed, while the genetic information recorded by DNA was permanently preserved and transmitted to ensure the continuation of the species.

ATP-CENTERED RNA WORLD DRIVEN BY SOLAR ENERGY

Woese [33] proposed the RNA world hypothesis (a term coined by Walter Gilbert in 1986 [34]): the earliest biomolecules on Earth were RNA, followed by DNA; the

early RNA molecules had the ability to store information as DNA and to perform catalysis as ribozymes and supported the operation of the early cell or protocell. The RNA world hypothesis is believed to be the most widely accepted hypothesis to explain how life arose [**35**]. Although there may never be direct physical evidence of an RNA-based organism, several lines of evidence are believed to support the RNA world hypothesis: discovery of catalytic ribozymes, catalytic roles of natural RNA, and evolution of enhanced catalytic function of RNA under *in vitro* selection [**17, 32, 36**].

Laboratory experiments on RNA evolution confirmed the roles of RNA in catalyzing nucleotide synthesis, RNA polymerization, aminoacylation of transfer RNA and peptide bond formation [**37, 38, 39, 40**]. For example, ribozymes can catalyze specific biochemical reactions, similar to protein enzymes. The *in vitro*-evolved ribozyme catalyzes the 5'-phosphorylation of polynucleotides using ATP-γ-S (or ATP) as a phosphate donor, with a rate enhancement of ~10^9-fold compared to the uncatalyzed reaction [**41, 42**].

It is not difficult to imagine that the catalytic functions of ribozymes perhaps played a transitional role in the early development of the biochemical system before their replacement by the more efficient protein enzymes. Proteins are more powerful in driving chemical reactions because of their diverse cationic, anionic and hydrophobic groups.

Interestingly, the nucleotide-derived coenzymes seemed to be remnants of an early RNA-based metabolism, as they still play a prominent role in most of these reactions today [**43**]. I believe that these factors do confirm the ability of early RNA (or more accurately NTP aggregates) in helping to build complex prebiotic systems. However, it is little persuasive evidence of the existence of organizing centers, analogous to modern ribosomes, where various RNAs and small molecules came together through noncovalent or transient covalent interactions [**44**].

Despite this, I still suspect the validity of the statement that RNA created the living world from randomness, although I agree with the view that RNA emerged earlier than DNA (**Figure 1**). Is there any evidence to say that the living world must have been derived from RNA as it has the functions of information storage and catalysis? Additionally, the RNA world hypothesis is unable to explain why RNA molecules tend to store genetic information and to support the operation of protocells.

In my opinion, the explanation of the RNA world hypothesis on the origin of life and the genetic code is somewhat farfetched [**45, 46, 47**], and its lack of both driving forces and individuality is unfortunate. Even if it turns out that RNA-based life once existed on the primitive Earth, this will not be a final answer, as one may ask where RNA originated [**31**]. We still would not know what happened before the emergence of RNA and why. Therefore, I propose here an alternative term, the ATP-centered RNA world (mediated by solar-photon-driven chemical synthesis), i.e., the early life on Earth coevolved with the development of the photosynthetic system in lipid vesicles where a sophisticated biochemical system was built to synthetize ATP using solar energy. This is evidenced by a series of structural and functional features of the extant biochemical system (e.g., photosynthetic pigments, electron transfer chain, ATP

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2020 doi:10.20944/preprints202003.0419.v5

synthesis by a transmembrane H^+ gradient, metabolic cycles/pathways coupled with ATP, and synthesis of RNA and DNA by ATP and its derivatives) as well as their interplay. However, if we are not willing to abandon the term "RNA world", then the nucleotide ATP should also be the key initiator.

PERSPECTIVES: WHERE TO GO?

The origin of life is summarized as "container-enabled chemistry first" or "self-assembling" [**48, 49**]. Perhaps some people will ask why molecules did not stay dormant in the organic soup and instead struggled to shift from chaos to order? We may attribute this to code-based self-organization with incessant input of solar energy: had such codes not existed, the protocells would have not been able to maintain orderly control of biochemical systems, and the chemical world have remained in incomprehensible chaos. However, an explanation is still necessary. Is it a first principle that individuals with heritable homogeneity were favored by existence (the state of being real)?

It is generally accepted that life requires structural complexity [**50**]. However, life, in a sense, represents a contradictory unity—it derives generality from homogeneity but also derives individuality from heterogeneity. Individuality can be traced back to the lipid vesicles where life started out. It is this quality that makes one living entity different from all others. It is amazing that such individuality developed from pure chemistry in protocells to give rise to sophisticated desire (e.g., for competition and struggle), habits, instincts and even spirit in higher animals. Is it possible that individuality, as an eco-physiological force, was engaged in reshaping or refixing the rhythm or regularity of biochemical reactions in protocells? Darwinian evolution could be the course that ultimately drove the formation of advanced forms of life.

One may well wonder whether too much speculation has been superimposed on the ATP hypothesis. Of course, all theories or hypotheses on the origin of the genetic code or life have been speculative, as billions of years have passed since their first appearance on Earth. However, a good theory/hypothesis depends on the ability to extensively interpret current biochemical systems. It should be kept in mind that the origin of life and its genetic codes are indivisible. Anyway, I argue that the present model is clearer than any other hypothesis in terms of the overall outline or logic, particularly in relating the origin of the genetic code with the evolution of the photochemical synthesis-mediated and ATP-centered biochemical system (Table 1). Although coevolution of the genetic code, protein synthesis and nucleicacid replication is discussed, no attention is paid to energetic forces [**51**]. It is, indeed, difficult to obtain experimentally testable model, as all the hypotheses or theories on the origin of life or genetic code can neither be verified nor falsified. This has been the case so far and perhaps will continue to be for a period of time. Particularly, the long parallel biochemical evolution in the primordial cell could be an important reason for this difficulty. However, because this is a question about ourselves, human beings will not stop pursuing the answer until the facts are completely known. It is possible for us to reveal the secrets of coevolution between the codons and the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2020 doi:10.20944/preprints202003.0419.v5

biochemical system based on their intrinsic relationships—if this does not necessarily reveal the truth, it may at least be a way to the truth.

Table 1 A comparison of the major theories on the origin of the genetic code

Name	Connotation	Evidences or inferences	Ref.
The stereo-chemical theory	There is, in many cases a specific stereochem-ical fit between an amino acid and the base sequence of the corresponding codon on the appropriate tRNA.	1. The code is universal because it is necessarily the way it is for stereochemical reasons.	[9]
The frozen accident theory	The allocation of codons to amino acids in a single ancestor occurred entirely by "chance" and then remained unchanged	1. The code is universal because at the present time any change would be lethal, or at least very strongly selected against. This accounts for the fact that the code does not change; 2. To account for it being the same in all organisms one must assume that all life evolved from a single organism.	[1]
The co-evolutio n theory	The code was assigned in parallel to the evolution of amino acid biosynthesis	1. The structure of the codon system is primarily an imprint of the prebiotic pathways of amino-acid formation, which remain recognizable in the enzymic pathways of amino-acid biosynthesis; 2. Consequently the evolution of the genetic code can be elucidated on the basis of the precursor-product relationships between amino acids in their biosynthesis; 3. The codon domains of most pairs of precursor-product amino acids should be contiguous, i.e., separated by only the minimum separation of a single base change.	[2]
The synthetical theory	The genetic code is a product of selection, history and chemistry	1. The pattern of codon assignments is an adaptation that optimizes some function, such as minimization of errors caused by mutation or mistranslation; 2. The genetic code accumulated amino acids over a long period of time and codon assignments reflect this pattern of incremental expansion; 3. Certain codon assignments may have been directly influenced by favorable	[10]

		chemical interactions between particular amino acids and short nucleic acid sequences, whereas lack of such interactions excluded other amino acids from proteins entirely.	
The ATP theory	The genetic code is a product of coevolution of ATP with the pristine biochemical system in the protocell, and ATP completed biochemical innovation from energy transformation to informatization through mediating synthesis of polynucleotides and polypeptides, leading to the emergence of a triplet codon system characterized by cyclizing of nucleic acid and proteins into a loop of reciprocal causation.	1. ATP is the only energetic product of photosynthesis translating solar energy to chemical energy, and is at the energetic heart of the extant biochemical systems; 2. Serving as both an energy carrier and an informatization molecule, ATP could energetically elongate chains of both polynucleotides and polypeptides, thus providing a bridge between these molecules and eventually mediating biochemical innovation in the protocell from energy transformation to informatization; 3. Informatization, a process for creating and managing information, was inevitably coupled with structuralization (processes for organizing or incorporating cellular structures) of the protocell, cyclizing polynucleotides and polypeptides into a feedback loop of reciprocal causation.	The present study

It is challenging to test the credibility of various theories/hypotheses proposed to explain the origin of the genetic code [52]. However, this is not a zero-sum game. The ATP hypothesis is new but is not in conflict with most of the existing dogmas, as the stereochemical, frozen accident, coevolution, synthetic and ambiguity reduction [53] hypotheses reflect each unique aspect of the story. The ATP hypothesis would greatly improve our understanding of the origin of the genetic code that handles information, a measurable abstract entity [54]. Nevertheless, the collection of supporting evidence to verify this hypothesis remains a challenge.

ACKNOWLEDGMENTS

The author thanks Professor Min S. Park for English improvements and insightful discussions. The author also thanks Drs. Gérard Battail, Vladimir Matveev, Andrew Moore, Martin B. Sweatman, Katsuhiko Ariga, Marcello Barbieri, Massimo Di Giulio and Joseph Moran for their critical but valuable comments and suggestions. This research was supported by the Strategic Priority Research Program of Chinese Academy of Sciences, grant no. XDB31000000.

Conflict of interest statement. None declared.

References

1. Crick, F.H. The origin of the genetic code. *J. Mol. Biol.* **1968**, 38, 367-379.
2. Wong, J.T.F. A co-evolution theory of the genetic code. *Proc Natl Acad Sci USA* **1975**, 72, 1909-1912.
3. Frank, A.; Froes, T. The standard genetic code can evolve from a two-letter GC Code without information loss or costly reassignments. *Orig. Life Evol. Biosph.* **2018**, 48, 259-272.
4. Vitas, M.; Dobovisek, A. In the beginning was a mutualism - on the origin of translation. *Orig. Life Evol. Biosph.* **2018**, 48, 223-243.
5. Vitas, M.; Dobovisek, A. Towards a General Definition of Life. *Orig. Life Evol. Biosph.* **2019**, 49, 77-88.
6. Rauchfuss, H. Chemical evolution and the origin of life. Springer-Verlag Berlin Heidelberg, 2008.
7. Yockey, H.P. Information theory, evolution, and the origin of life. Cambridge University Press: Cambridge, UK, 2005.
8. Gamow, G. Possible relation between deoxyribonucleic acid and protein structures. *Nature*, **1954**, 173: 318.
9. Woese, C.R.; Dugre, D.H.; Dugre, S.A.; Kondo, M.; Saxinger, W.C. On the fundamental nature and evolution of the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* **1966**, 31, 723-736.
10. Knight, R.D.; Freeland, S.J.; Landweber, L.F. Selection, history and chemistry: the three faces of the genetic code. *Trends Biochem. Sci.* **1999**, 24, 241-247.
11. Baranov, P.V.; Atkins, J.F.; Yordanova, M.M. Augmented genetic decoding: global, local and temporal alterations of decoding processes and codon meaning. *Nature Rev. Genet.* **2015**, 16, 517-529.
12. Koonin, E.V.; Novozhilov, A.S. Origin and evolution of the genetic code: the universal enigma. *Cell Mol. Biol.* **2009**, 61, 99-111.
13. Ohama, T.; Inagaki, Y.; Bessho, Y.; Osawa, S. Evolving genetic code. *Proc. Jpn. Acad. Ser. B.* **2008**, 84, 58–74.
14. Shah, P,; Gilchrist, M.A. Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc. Natl. Acad. Sci. USA* **2011**, 108, 10231-10236.
15. Sciarrino, A.; Sorba, P. Codon-anticodon interaction and the genetic code evolution. *Biosystems* **201**3, 111, 175-180.
16. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **1971**, 58, 465-523.
17. Joyce, G.F. The antiquity of RNA-based evolution. *Nature* **2002**, 418. 214-221.
18. Leslie, M. On the origin of photosynthesis. *Science* **2009**, 323, 1286-1287.
19. Umena, Y.; Kawakami, K.; Shen, J.R.; Kamiya, N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* **2011**, 473, 55-60.

20. Nelson, D.L.; Cox, M.M. Lehninger Principles of Biochemistry (4th ed.). W.H. Freeman, 2004..
21. Battail, G. Information and Life. Springer, 2014.
22. Michaelian, K.; Simeonov, A. Fundamental molecules of life are pigments which arose and evolved to dissipate the solar spectrum. *Biogeosciences Discuss*. **2015**,12, 2101-2160.
23. Xie, P. The Aufhebung and Breakthrough of the Theories on the Origin and Evolution of Life. Science Press: Beijing, P.R. China, 2014.
24. Taiz, L.; Zeiger, E. Plant Physiology (4th edition). Sinauer Associates, 2010.
25. Madigan, M.T.; Martinko, J.M.; Stahl, D.A.; Clark, D.P. Brock Biology of Microorganisms (13th ed). Prentice Hall, 2012.
26. Michaelian, K. Thermodynamic dissipation theory for the origin of life. *Earth Syst. Dynam.* **2011**, 2, 37-51.
27. Powner MW, Gerland B, Sutherland JD. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature **2009**, 459, 239-42.
28. Patel BH, Percivalle C, Ritson DJ, Duffy CD, Sutherland JD. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. Nat Chem. **2015**, 7, 301-307.
29. Wong, Tze Fei JWong JT, Lazcano A. Prebiotic Evolution and Astrobiology. Landes Bioscience, Austin Texas, 2009.
30. Wong, J.T.F.; Lazcano. A. (eds.) Prebiotic Evolution and Astrobiology. CRC Press, 2009.
31. Orgel, L.E. Evolution of the genetic apparatus: a review. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, 52, 9-16.
32. Doudna, J.A.; Cech, T.R. The chemical repertoire of natural ribozymes. *Nature* **2002**, 418, 222-228.
33. Woese, C.R. The Genetic Code: the Molecular Basis for Genetic Expression. Harper & Row, 1967.
34. Gilbert, W. The RNA world. *Nature* **1986**, 319, 618.
35. Griffith, R.W. A specific scenario for the origin of life and the genetic code based on peptide/oligonucleotide interdependence. *Orig. Life Evol. Biosph.* **2009**, 39, 517-531.
36. Cech, T.R.; Zaug, A.J.; Grabowski, P,J, *In vitro* splicing of the ribosomal RNA precursor of *Tetrahymena*: Involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **1981**, 27, 487-496.
37. Zhang, B.; Cech, T.R. Peptide bond formation by *in vitro* selected ribozymes. *Nature* **1997**, 390, 96-100.
38. Unrau, P.J.; Bartel, D.P. RNA-catalysed nucleotide synthesis. *Nature* **1998**, 395, 260-263.
39. Lee, N.; Bessho, Y.; Wei, K.; Szostak, J. W.; Suga, H. Ribozyme-catalyzed tRNA aminoacylation. *Nature Struct. Biol.* **2000**, 7, 28-33.
40. Johnston, W.K.; Unrau, P.J.; Lawrence, M.S.; Glasner, M.E.; Bartel, D.P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **2001**, 292, 1319-1325.

41. Lorsch, J.; Szostak, J.W. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **1994**, 371, 31-36.
42. Huang, F.; Yarus, M. Versatile 5' phosphoryl coupling of small and large molecules to an RNA. *Proc. Natl Acad. Sci. USA* **1997**, 94, 8965-8969.
43. White, H.B. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **1976**, 7, 101-104.
44. Gibson, T.J.; Lamond, A.I. Metabolic complexity in the RNA world and implications for the origin of protein synthesis. *J. Mol. Evol.* **1990**, 30, 7-15.
45. Atkins, J.F.; Gesteland, R.F.; Cech, T.R. (eds.) RNA worlds: From Life's Origins to Diversity in Gene Regulation. Cold Spring Harbor Laboratory Press, 2011.
46. Sengupta, S.; Higgs, P.G. Pathways of genetic code evolution in ancient and modern organisms. *J. Mol. Evol.* **2015**, 80, 229–243.
47. Zimmer, C. On the origin of life on earth. *Science* **2009**, 323, 198-199.
48. Deamer, D. Assembling Life: How can Life Begin on Earth and other Habitable Planets? Oxford University Press: New York, NY, USA, 2019.
49. Bains, W. Getting beyond the toy domain. Meditations on David Deamer's "Assembling Life". Life 2020, 10, 18; https://doi.org/10.3390/life10020018.
50. Mayer, C. Life in the context of order and complexity. *Life* **2020**, 10, 5, https://doi.org/10.3390/life10010005.
51. Alberti S. Evolution of the genetic code, protein synthesis and nucleic acid replication. *Cell. Mol. Life Sci.* **1999**, 56, 85-93.
52. Di Giulio, M. A discriminative test among the different theories proposed to explain the origin of the genetic code: The coevolution theory finds additional support. *BioSystems* **2018**, 169-170, 1-4.
53. Barbieri, M. Evolution of the genetic code: the ambiguity-reduction theory. *BioSystems* **2019**, 185, 104024.
54. Battail, G. Error-correcting codes and information in biology. *BioSystems* **2019**, 184, 103987.