 doi:10.20944/preprints202003.0419.v2

ATP-hypothesis: a new conceptual framework for the origins of genetic codes

Ping Xie[1, 2]

[1] University of Chinese Academy of Sciences, Beijing 100049, China

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

* Correspondence: xieping@ihb.ac.cn

Abstract

The origin of genetic codes is the key to reveal life's origin on the earth as it is a prerequisite for the existence of life. More than half a century has passed since the discovery of genetic codes, while their origin is still one of the greatest mysteries. Are the origins of genetic codes really unknowable? Do they really require external design? Here, I present an ATP-hypothesis that explains how the genetic codes came into being with the coevolution of biochemical system. ATP has several properties that make it suitable as the initiator of the origin of genetic codes. First, ATP is the only energetic product of photosynthesis. Second, ATP is at the heart of the extant biochemical systems. Third, ATP serves as carriers of both energy and information. Fourth, ATP could energetically elongate chains of both polynucleotides and polypeptides, thus providing a bridge between them, and eventually mediating prebiotic biochemical transaction from energy to information. This hypothesis shows how primitive life emerged through a series of processes from energy to information mediated by ATP. Informatization (processes of creating and managing information) was inevitably coupled with structuralization (processes of organizing or incorporating into a cellular structure), making polynucleotides and polypeptides be cyclized into a system of reciprocal causation. The triplet codon might just be for stereo-chemical handing of amino acid through e.g. Watson–Crick pairing interactions. It is an evolutionary completion for genetic codes from RNA to DNA, only which, a reverse to the Central Dogma, marked the dawn of cellular life when Darwinian evolution began to operate. ATP-hypothesis shades lights on the origin of life, together with the formations of both photosynthesis and biochemical systems, which have been largely unknown so far.

Key words: origin of genetic codes, ATP-hypothesis, solar energy driven synthesis of chemicals, informatization, structuralization

1. Introduction

Genetic information is molecular memory recorded by triplet codon, 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 the earth. But no one knows exactly how these genetic codes came into being. Evolution of genetic code can be discussed within the context of the evolution of biochemical reactions in the protocell where ATP was at the center and became the key first building block which initiated a series of potential chain reactions that resulted in triplet codons.

1-1. Agnosticism

Biologists, probably as the majority, hold a pessimistic view that 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 trace in physics or chemistry of the control of chemical reactions by a sequence of any sort or of a code between sequences. He criticizes that many papers have been published with titles indicating that their subject is the origin of the genetic code, but actually the content deals only with its evolution.

1-2. Prevalent hypotheses and their deficiencies

So far, several hypotheses have been proposed. The frozen accident hypothesis states that allocation of codons to amino acids in the single ancestor was entirely a matter of "chance", and then remained unchanged [1]. The stereochemical hypothesis claims that there is in many cases a specific stereochemical fit between the amino acids and the base sequence of its codon on the appropriate tRNA[8, 9]. The biosynthetic hypothesis postulates that the code was assigned in parallel to the evolution of amino-acid biosynthesis [2]. Knight et al. [10] declares that the genetic code is a product of selection, history and chemistry. However, none of these describe how genetic codes were originally created. Since then, very little definitive progress has been made, although the literature is replete with attempts to explain variation or flexibility of the codes and possible rules of codon allocations to amino acids [11, 12, 13, 14].

Frankly speaking, these hypotheses suffer from two defects: first, none can explain satisfactorily why the genetic codes evolved in such a way, and second, none has explained the origin of genetic codes from that of the biochemical system (a relation of part to whole). In other words, all these hypotheses completely overlooked the coevolution of the genetic codes with biochemical system. In my view, it is impossible to crack the secret of codon origin just from the codon itself [11, 15], even extending vision to the possible relationship between codon and amino acids [13].

1-3. Chicken-and-egg paradox

Another source of the problem may be related to a so called quasi-species model proposed by Eigen [16]. He strongly advocated *in vitro* evolution of macromolecules.

 doi:10.20944/preprints202003.0419.v2

A quasi-species in the environment was imaged to be a population of genetically related RNA molecules which had certain morphological commonness but were not identical. He supposed that the quasi-species followed the Darwinian process of natural selection. This model has been highly influential [6], which, in my view, is perhaps misleading our understanding the origin of the genetic codes. For example, it is erroneously assumed that there was a random succession of dominant polymer sequences and their associated functional properties in natural history of life on primitive earth [17]. It should be born in mind that all extant cells are surrounded by a membrane composed of amphipathic lipids.

Molecular biology has been long troubled by the so-called chicken-and-egg paradox of protein and nucleic acids, seemingly a logically circular debate about who 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 cyclizing of biochemical paths where the transitional states or tracks had long disappeared. Just as the ancient Greek philosopher Heraclitus said, the beginning and ending point are overlapped on the circumference of a circle. It is great to innovate protein synthesis handled by RNA, and vice versa.

2. ATP-hypothesis

In this article, I present a new hypothesis – ATP hypothesis to explain the central role of ATP in the origin of genetic codes (**Figure 1**) because 1) ATP was the first building block and was renewed by solar-photon driven chemical reactions in the protocell, 2) ATP has key features making it an energy carrier and an informatization molecule, 3) ATP connected amino acids into short peptides and catalytic polypeptide, 4) ATP drove the formation of nucleotides which resulted in thermodynamically favorable and/or probabilistic selection of triplet codon, 5) RNA served as the carrier of information and, and 6) DNA as the final storage molecule of genetic information by housing the genetic codes. As ATP provides energy for the synthesis of all other nucleotides, while AGCT/U comprise the genetic codes, in this sense, ATP is a carrier of information.

Key arguments favoring ATP as the initiator of the origin of genetic codes are its ability to elongate chains of both polynucleotides and polypeptides without additional energy input, which made 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 a feedback mechanism (selection of cellular survival); and technically, photosynthesis, a goal-oriented process, enabled various biotic factors or reactions (ATP, lipid vesicle, informatization, structuralization, homogenous individual, individuality, survival, etc.) to be integrated into an operating system of genetic codes.

It is undeniably important but largely conjectural issue how life with its genetic code came into being on the primitive earth more than three billion years ago. Nevertheless, I can make a logical conjecture in light of philosophical perspectives. The ancient Greek philosopher Aristotle once said, ‘we do not have knowledge of a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2020 doi:10.20944/preprints202003.0419.v2

thing until we have grasped its why, that is to say, its cause’. Thus, if in terms of Aristotelian four causes, “matter” was composed of various building blocks (e.g., nucleotides, amino acids, sugars, lipids etc.), “form” makes matter into a particular type of thing - protocells with biochemical systems including genetic codes, biochemical cycles, etc., “agent” were ATP produced by photo-chemical reactions to cause change, and “purpose” is goal of the form, i.e., individuality of the cell and its desire to struggle for survival.

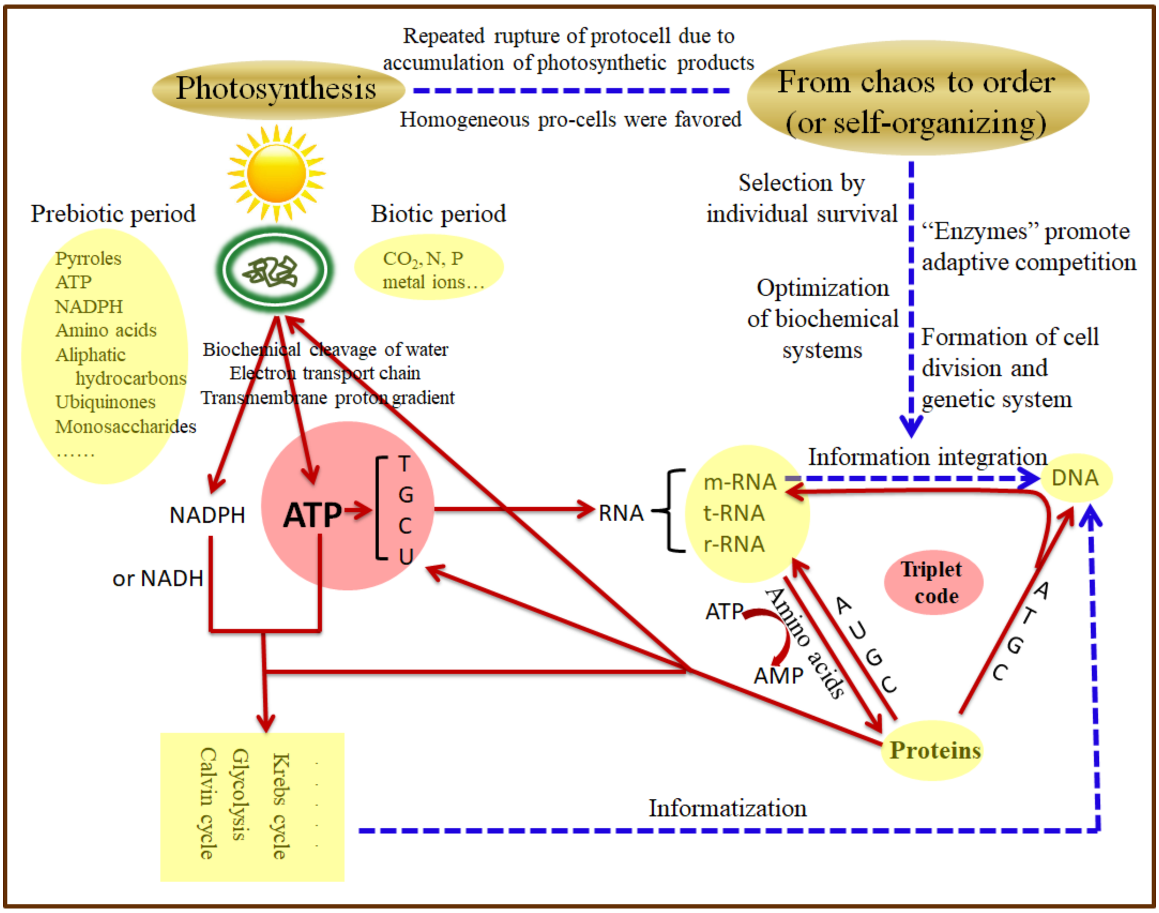

Figure 1. A simplified conceptual model on the origin of the genetic code based on the photosynthesis-mediated and ATP-centric hypothesis. Dashed blue lines indicate evolutionary processes during pre-life period, while solid red lines denote processes or interactions from pre-life period to the present. Arrows indicate the direction of influences or actions. In the prebiotic period, there existed a plenty of life's building blocks (e.g., pyrroles, ATP, NADPH, amino acids, aliphatic hydrocarbons, ubiquinones, monosacharides etc.). Impermeable lipid vesicle, life's building block, and sunlight are basis for the origin of life. The biotic period is characterized by the establishment of the genetic code system by a photochemical mechanism related closely with the production of ATP (e.g. biochemical cleavage of water, electron transport chain, transmembrance proton gradient). ATP is at the heart of the extant biochemical systems (e.g. Krebs cycle, glycolysis, Calvin cycle), and is a bridge between polynucleotide and polypeptide, as it serves as carriers of both energy and information. Informatization was coupled with cellular structuralization, making polynucleotides and polypeptides becyclized into a

system of reciprocal causation, then with the emergence of genetic codes. It is an evolutionary completion for genetic codes from RNA to DNA, only which, a reverse to the Central Dogma, marked the dawn of cellular life when Darwinian evolution began to operate.

3. Discussion

3-1. Why ATP?

"Hidden hand" or natural forces?

It is undeniable that codon and amino acids had been linked stereo-chemically [9]. Otherwise, we will fall into the mire of 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 the ones, but I do believe they are rather than driving forces. I contend that driving forces should be energetic, e.g. ATP produced by solar-photon driven synthesis of chemicals in primordial cell.

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

Inspiration from the organization of extant biochemical system

To solve the puzzle of the genetic codes, we must first understand how a biochemical system is organized. Consider what is life. Life, compositionally, is a unity of matter, energy and information, and dynamically, is a harmonious play of material cycling, energy flows and information communication. Energy is the key to support life system. In the past decades, physicists and chemists have discovered a lot of details about how organisms are structured and how they work. It is known that life is a physiological machine where inconceivably numerous biochemical reactions are taking place to acquire, convert and use energy. In all modern autotrophic life forms (plants, algae and some bacteria), the only source of energy is from sunlight that is converted into chemical energy by a series of complex physico-chemical processes called as photosynthesis [19].

Energetically and informatically, nothing in biochemical system is more important than a nucleotide, ATP. It is the only energetic product of photosynthesis, carrying chemical energy converted from sunlight. It then provides energy for metabolisms through conversions of ATP/ADP/AMP, supporting 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 the Krebs) are all coupled with ATP (**Figure 1**). Of course, NAD(P)H, a derivative of nucleotide, is also necessary, as it transports H and e^- (through conversion of $NAD(P)H/NAD(P)^+$).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2020 doi:10.20944/preprints202003.0419.v2

In extant life, ATP is not only an indispensable building block of the genetic system (DNA, and RNA), but the other four nucleotides for genetic coding are also all derived from it with some other life's building blocks, 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. While, information delineates the border between the living and the inanimate, as the living world appears as the only place where information is recorded, processed, or used [21]. Therefore, as an irreplaceable carrier of both energy and information, ATP appears to play a central role in the biochemical system.

Cues from 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 [23]. First, sunlight, needless to say, has been the most universal source of energy. Second, a biomolecule called cytochrome (an electron transport protein with iron porphyrin or heme as a prosthetic group) seems to be imprinted with photosynthesis. Cytochrome, a life's building block in "prebiotic soup", is an important part of the primitive photosynthetic system and thus is necessary for the production of ATP. Cytochrome is a universal electronic carrier, but is present even in chemoautotrophic bacteria [24]. Originally, the heme was likely derived from a photosynthetic pigment, chlorophyll, as their biosynthetic pathways are very similar in extant life forms [25]. Interestingly, chlorophyll seemed an adduct between a magnesium porphyrin ring and a long chain fatty acid from the membrane. Let me ask, if photosynthetic bacteria were not the Last Universal Common Ancestor (LUCA) of all modern life forms, why had the chemoautotrophic bacteria used a photosynthesis-imprinted molecule like cytochrome as an electron carrier?

The thermodynamic dissipation theory for the origin of life supposes 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 which arose and evolved to dissipate the solar spectrum [22]. The present ATP- hypothesis further details how life began with the evolution of biochemical system driven by photo-chemical synthesis in which ATP played a key role in the transformation biochemically from energy to information, leading to the birth of genetic codes.

The present paper rests on the predicate that life arose via photochemical synthesis, i.e. a primitive kind of photosynthesis was the origin of life, although the most plausible current theory on the origin of life is that which concerns the chemistry of alkaline thermal vents in the sea floor.

3-2. How ATP?

ATP mediated transaction from energy to information

The genetic code system was built by a chemical mechanism related closely with the production of ATP. Energetically, the first job the primordial life did was perhaps how to achieve sufficient production of this nucleotide. While in present-day photoautotrophs, synthesis of ATP requires a transmembrane gradient of protons that come from biochemical cleavage of water in photosynthetic system. Therefore, to guarantee such a proton gradient, there was first needed a relatively closed entity or a small room, impermeable to H^+. Such cubicles were most likely a lipid vesicle, the precursor of 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.

Chemically, it was not impossible, on the primitive earth, that fatty acids could automatically form a double-layered globular membrane structure [27]. In modern life forms, cell membrane, consisting of lipid (phospholipid) bilayer, provides controlled entry and exit ports for the exchange of matter. It permits passing through 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, unless using transmembrane protein channels. This implies that the formation of ATP in primordial cell gradually began to rely on polypeptide channel that then developed into ATP synthase. In addition, H^+ from cleavage of H_2O also needed the help of polypeptides.

Reasonably, 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……) had been organized orderly. It may be inferred that there were a series of events randomly occurring in protocells. For example, the transmembrane proton gradient coupled with polypeptide channel resulted in the formation of an apparatus (i.e. ATP synthase) to synthetize ATP. Nucleotides like ATP could form polynucleotides by self-condensation, some of which carried amino acids (precursor of tRNA), and others built 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 can carry out all reactions of protein synthesis. On the other hands, 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 self-condensation of nucleotides as well. As a consequence, numerous consecutive reactions were linked into a variety of chains, which might be linear, branched, or cyclic. Fundamentally, these were creative processes of order out of disorder, and of rationality out of randomness. In short, by these processes, the randomly existing life's building blocks were finally organized into ordered cellular structure and functions.

Primitive life commenced its journey from materials that were not organized. The prebiotic mixtures of chemicals (also called as "prebiotic soup"), present on the primeval earth, were assumed to contain an immense number of life's building blocks such as ATP, NADPH, pyrroles, amino acids, aliphatic hydrocarbons, ubiquinones, monosaccharides, and so on[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, catalysis [17].

Let us first consider the impermeable lipid vesicle (enclosed by amphiphiles having both polar and non-polar domains) wrapping a plenty of life's building blocks. Their exposure to sunlight would drive dazzling flows of electrons and H^+, causing active re-combinations of elements. This included various organic chemical reactions, leading to a series of prebiotic organic synthesis. This would increase accumulation of large molecules, but accompanied with ceaseless input of small molecules like CO_2……and therefore, the protocells had to reciprocate between enlarging and rupturing, giving rise to the developments of both photochemical synthesis (precursor of photosynthesis here called as quasi-photosynthesis) and cell division. Next, an information era followed: in the cycle of quasi-photosynthetic growth followed by division, and also through the selection of individual survival, the protocell established regularity, reproducibility and rhythmicity of various organic chemical reactions in dealing with photo-chemically synthetized products, consequently leading to a complex web of biochemical cycles and photosynthesis, respectively. This is a process to establish information (called as informatization), with cellular compartmentalization as well. Thus, such "reverse micelles" enclosing life's building blocks was likely the site where life with its genetic codes began to arise, driven by photo-chemical reactions.

The way 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 can not be excluded, but this was not favored by prebiotic selection. A scenario of how a set of genetic codes were successfully selected to record, preserve and transmit information is outlined below.

① ATP is a bridge between polynucleotide and polypeptide

Firstly, 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 into 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 nucleic acids (DNA/RNA) and proteins, then making it possible to create their relation, that is the information! In this way, ATP mediated biochemical transaction from energy to information. This is an indispensable step for the births of genetic codes and life on the earth, and also the key to the start to build biochemical system including genetic system in the protocell.

② From informatization to structuralization

Secondly, informatization was inevitably coupled with structuralization, such as structural subdivision or specialization and functional differentiation, providing 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 acid, polypeptides helped matching the acceptor stem of t-RNA 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 m-RNA template), and so on. In this way, macromolecules became functionally differentiated, i.e., handling (record, preserve and transmit) information by polynucleotides and catalyzing all chemical reactions by polypeptides called as enzymes, and both were further cyclized into a system of reciprocal causation. As a result, the code-based informatization led to evolutionary innovations of diverse principles or patterns, e.g., biochemical pathways (**Figure 1**). These prebiotic chemical processes substantially increased the complexity of the protocell. Foundation of a hereditable bio-information system marked a real shift of the world from prebiotic chemistry to primitive biology, and only after this stage, Darwinian evolution began to operate. Progressive refinements of these mechanisms then provided further selective advantages for protocells.

③ Why is triple?

Thirdly, 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 acid, it was not so good to be either too many (more cumbersome) or too few (4^3 =64, thus there is still considerable redundancy of encoding). Namely, triplet was the lowest code number to encode the amino acids.

④ Why hereditable homogeneity was favored?

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

It is suggested that only macromolecules that operated on their environment to produce further copies of themselves had an evolutionary future [**28**]. If in terms of individual (protocell), this seems a plausible view.

⑤ From RNA to DNA

It is a great mystery how nucleic acid diverged into DNA and RNA. The primary advantage of DNA over RNA as a genetic material is assumed to be the greater chemical stability of DNA, allowing much larger genomes based on DNA [17], although there are only minor structural differentiations.

As ATP is the only nucleotide produced by photosynthesis, it did not appear difficult to convert ATP to any other nucleotides. Meanwhile, these energetic nucleotides could also self-condense into a wide variety of mRNA, of course also other RNAs. If they 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 the informatization in biochemical system. That is, DNA was specialized to accurately record and permanently preserve all genetic information that command all that goes in the cell, while mRNA became short-lived messengers for the implementation of the instructions. It is an astonishing imprint for the evolutional way from RNA to DNA that RNA is still involved as a primer in DNA replication of extant organisms.

In the integrative process of information, there were subtle structural differences between DNA and RNA. First, the second carbon atom of the ribose connected -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 can answer why base differences were 1 rather than 4. This is perhaps just because of a need in difference, as similar cases were not rare, e.g. NADPH and NADH. Country to DNA, RNA can folds into a variety of complex tertiary structures, analogous to structured proteins, and catalyses a broad range of chemical transformations [29]. It is uncertain whether the diverse three-dimensional structures of RNA were due only to such a tiny change in ribose or base or not. 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 a tiny cell where hundreds of biochemical reactions occur simultaneously. Consequently, the biochemical system came to a status where mRNA were immediately destroyed after their mission was finished, while the genetic information recorded by DNA would be permanently preserved and transmitted to ensure the continuation of the species.

3-3. ATP-centered RNA world driven by solar energy

Woese [30] proposed the RNA world hypothesis (a term coined by Walter Gilbert in 1986[31]): the earliest biomolecules on earth were RNA, and then DNA (**Figure 1**); the early RNA molecules had the abilities of information storage as DNA and catalysis as ribozymes, and supported the operation of the early cell or the protocell life. The RNA world is believed to be the most widely accepted hypothesis to explain how life arose [32]. Although there may never be direct physical evidence of an RNA-based organism, several evidences were believed to support the RNA world: discovery of catalytic ribozymes, catalytic roles of natural RNA, and evolution of enhanced catalytic function of RNA under *in vitro* selection [17, 29, 33].

Laboratory experiments on RNA evolution confirmed the roles of RNAs in catalyzing nucleotide synthesis, RNA polymerization, aminoacylation of transfer

RNA and peptide bond formation [34, 35, 36, 37]. For example, ribozymes can catalyze specific biochemical reactions, similar to protein enzymes. The *in vitro*-evolved ribozyme catalyzes 5'-phosphorylation of polynucleotides using ATP-γ-S (or ATP) as phosphate donor, with a rate enhancement of ~10^9-fold compared to the uncatalyzed reaction [38, 39].

It is not difficult to imagine that the catalytic functions of ribozymes perhaps played a transitional role in the early development of 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 remnants of an earlier RNA-based metabolism, as they still played a prominent role in most of these reactions today [40]. I believe that these do confirm technical ability of early RNA (or more accurately NTP aggregates) in helping to build complex prebiotic systems. However, it is little persuasive to say that there were organizing centres, analogous to modern ribosomes, where various RNAs and small molecules came together through non-covalent or transient covalent interactions [41].

Despite all this, I still suspect validity of such a statement that RNA created the life world from randomness, although agreeing with the view that RNA was earlier than DNA. Is there any evidence to say that the world of life must have been derived from RNA as it has the functions of information storage and catalysis? Also, RNA world hypothesis is unable to explain why the RNA molecules tended to store genetic information and to support the operation of protocells.

In my view, the explanation of the RNA world hypothesis on the origin of life and the genetic codes is somewhat farfetched [42, 43, 44] 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, it will not be a final solution, as one may ask where RNA came from [28]. We may still do not know what happen before this and why. Therefore, I propose here an alternative term, ATP-centered RNA world (mediated by the solar-photon driven chemical synthesis), i.e., the early life on earth coevolved with the development of 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 synthesis by transmembrane H^+ gradient, metabolic cycles/paths coupled with ATP, synthesis of RNA and DNA by ATP and its derivatives, etc.) as well as their interplay. However, if we are not willing to abandon the term of "RNA world", well, the nucleotide ATP should also be the key initiator.

It should be noted, however, that there is not yet a general consensus in the origins of life community about whether the earliest organisms were photosynthetic or chemosynthetic or heterotrophic. ATP-hypothesis also strongly supports photosynthetic origin of life.

4. Perspectives: Where to go?

The origin of life is summarized as "container-enabled chemistry first" or "self-assembling" [45, 46]. Perhaps some people will ask why molecules did not stay silently in the organic soup but struggled to shift from chaos to order? We may attribute this to code-based self-organization with incessant input of solar energy: would such codes not exist, the protocells could have not been able to maintain orderly control of biochemical systems, and the chemical world would still stay in incomprehensible chaos. But, it is still needed to explain why. 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 [47]. However, life, in a sense, is a contradictory unity – it owns generality from homogeneity, but in the meantime possesses individuality from heterogeneity. Individuality can be traced back to the lipid vesicles where life started out. It is the quality that makes one living entity different from all others. It is amazing that such individuality had advanced from pure chemistry in protrocells 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, engaged in re-shaping or re-fixing rhythm or regularity of biochemical reactions in protocells? And then, Darwinian evolution could be a matter of course that ultimately drove the formation of advanced form 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 genetic codes or life have been speculative, as billions of years had passed since their first appearance on earth. However, a good theory/hypothesis depends on the ability to have extensive interpretation on current biochemical systems. It should be kept in mind that origins of life and its genetic codes are indivisible. Anyway, I argue that the present sketch is clearer than any other hypotheses in the overall outline or logic, particularly, relating the origin of the genetic code with the evolution of the photochemical synthesis-mediated and ATP centered biochemical system. And, of course, I admit that 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 still be for a period of time in the future. But, because this is a secret about ourselves, human beings never stop pursuing before the facts are completely known. It is possible for us to spy on the secrets of coevolution between the codon and the biochemical system from their intrinsic relationships – if this is not necessarily the truth, but at least might be a way to the truth.

It is challenging to test credibility of various theories/hypotheses proposed to explain the origin of the genetic code [48]. However, this is not a zero-sum game. ATP-hypothesis is new but not conflicting with most of the existing dogmas, as stereochemical, frozen accident, coevolution, synthetic and ambiguity-reduction [49] hypotheses respectively reflect each unique profile of the story. ATP-hypothesis would greatly improve our understanding the origin of genetic codes that handle information, a measurable abstract entity [50]. Nevertheless, it still remains a challenge to collect more supporting evidences to justify this hypothesis.

Declaration of Competing Interest

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

Acknowledgement

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

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

 doi:10.20944/preprints202003.0419.v2

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 (4^{th} 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. Madigan, M.T.; Martinko, J.M.; Stahl, D.A.; Clark, D.P. Brock Biology of Microorganisms (13th ed). Prentice Hall, 2012.
25. Taiz, L.; Zeiger, E. Plant Physiology (4th edition). Sinauer Associates, 2010.
26. Michaelian, K. Thermodynamic dissipation theory for the origin of life. *Earth Syst. Dynam.* **2011**, 2, 37-51.
27. Wong, J.T.F.; Lazcano. A. (eds.) Prebiotic Evolution and Astrobiology. CRC Press, 2009.
28. Orgel, L.E. Evolution of the genetic apparatus: a review. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, 52, 9-16.
29. Doudna, J.A.; Cech, T.R. The chemical repertoire of natural ribozymes. *Nature* **2002**, 418, 222-228.
30. Woese, C.R. The Genetic Code: the Molecular Basis for Genetic Expression. Harper & Row, 1967.
31. Gilbert, W. The RNA world. *Nature* **1986**, 319, 618.
32. 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.
33. 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.
34. Zhang, B.; Cech, T.R. Peptide bond formation by *in vitro* selected ribozymes. *Nature* **1997**, 390, 96-100.
35. Unrau, P.J.; Bartel, D.P. RNA-catalysed nucleotide synthesis. *Nature* **1998**, 395, 260-263.
36. Lee, N.; Bessho, Y.; Wei, K.; Szostak, J. W.; Suga, H. Ribozyme-catalyzed tRNA aminoacylation. *Nature Struct. Biol.* **2000**, 7, 28-33.

37. 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.
38. Lorsch, J.; Szostak, J.W. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **1994**, 371, 31-36.
39. 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.
40. White, H.B. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **1976**, 7, 101-104.
41. 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.
42. 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.
43. Sengupta, S.; Higgs, P.G. Pathways of genetic code evolution in ancient and modern organisms. *J. Mol. Evol.* **2015**, 80, 229–243.
44. Zimmer, C. On the origin of life on earth. *Science* **2009**, 323, 198-199.
45. Deamer, D. Assembling Life: How can Life Begin on Earth and other Habitable Planets? Oxford University Press: New York, NY, USA, 2019.
46. Bains, W. Getting beyond the toy domain. Meditations on David Deamer's "Assembling Life". Life 2020, 10, 18; https://doi.org/10.3390/life10020018.
47. Mayer, C. Life in the context of order and complexity. *Life* **2020**, 10, 5, https://doi.org/10.3390/life10010005.
48. 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.
49. Barbieri, M. Evolution of the genetic code: the ambiguity-reduction theory. *BioSystems* **2019**, 185, 104024.
50. Battail, G. Error-correcting codes and information in biology. *BioSystems* **2019**, 184, 103987.