 doi:10.20944/preprints202003.0419.v8

HYPOTHESIS

The ATP hypothesis discovers the missing "matchmaker" between proteins and nucleic acids

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 the energetic driving force, its relation to the biochemical system, and most importantly, the missing "matchmaker" between proteins and nucleic acids. Here, a new hypothesis is proposed, according to which ATP is at the origin of the primordial genetic code by driving the coevolution of the genetic code 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 vesicle 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. ATP could energetically form and elongate chains of both polynucleotides and polypeptides, thus acting a "matchmaker" between these two bio-polymers and eventually mediating precellular biochemical innovation 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 a photosynthesis system to produce ATP efficiently and regularly with the aids of proteins and RNA/DNA. The completion of permanently recording the genetic information by DNA marked the dawn of cellular life operated by Darwinian evolution. The ATP hypothesis supports the photochemical origin of life, shedding light on the origins of both photosynthetic and biochemical systems, which remains largely unknown thus far.

Key words: ATP hypothesis, origin of genetic code, life's building block, probiotic "soup", coevolution, biochemical system, missing "matchmaker", energy transformation, informatization, structuralization, precellular selection, photo-chemical origin of life

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

Contents

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

1. Background

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, as it is shared by almost every kind of organism, from bacteria to humans, with only a few minor exceptions, implying that the standard code was used by a common ancestor of all forms of life in its actual form within a very short period of time at or very near the supposed beginning of life on Earth [6]. Therefore, the code is the universal language of life.

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? We are too much far away from reaching any clear concept on how such genetic codes could have evolved some 3.5 billion years ago. Or it is still unclear why the genetic code might have originated in the prebiotic world leading to the *information* age [7]. Cavalier-Smith [8] says, "Only selection at a higher level than for individual selfish genes could power the cooperative macromolecular coevolution required for evolving the genetic code". I totally agree with this view. Therefore, it is a wrong view that life began by the origin of replication by naked genes and that cells evolved later. In principle, only cellular selection could increase complex properties like the genetic code.

In this paper, I present the ATP hypothesis as a new conceptual framework for the origin of the genetic code. Evolution of the genetic code is discussed within the context of the evolution of biochemical reactions in the lipid vesile and then protocell where ATP was at the bioenergetic center to transform solar energy into chemical energy, initiating a series of chain reactions, yielding a triplet codon to record information.

1-1. Evolution≠Origin

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 [9]. Smith and Szathmáry [10] wrote: "the origin of the code represents the most perplexing problem in evolutionary biology; the existing translational machinery is at the same time so complex, so universal, and so essential that it is hard to see how it could have come into existence, or how life could have existed without it". Yockey [11] 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

1-2. Fatal deficiencies of previous theories

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 [12, 13]. The stereochemical and frozen accident hypotheses are seemingly opposing views. It is also argued that the code is neither an accident nor totally frozen [8]. The biosynthetic hypothesis postulates that the code was assigned in parallel to the evolution of amino acid biosynthesis [2]. Knight et al. [14] declared that the genetic code is a product of selection, history and chemistry. However, in my view, 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 [15, 16, 17, 18].

Frankly speaking, these hypotheses suffer from several defects: first, all overlooked the importance of the energetic driving force, 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). Especially, these hypotheses also ignored the amazing relation between energy transformation and informatization. In my opinion, it is impossible to understand the origin of codons based on the codons alone [15, 19] or even by extending the perspective to the possible relationship between codons and amino acids [17].

1-3. Molecules never evolved *in vitro*

In my view, evolution of a protocell must have started from a confined environment (e.g. lipid vesicle). Actually, we have long been troubled by the so-called quasispecies model proposed by Eigen who strongly advocated for the *in vitro* evolution of macromolecules [20]. 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 [9], 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 [21]. 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. It is well known that amphipathic molecules with long hydrophobic tails and hydrophilic heads can spontaneously form membranes in the presence of liquid water [8].

1-4. *In vivo* coevolution of the bio-polymers

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

It is great for a cell to innovate protein synthesis manipulated by RNA, and vice versa. This is the so-called "chicken-and-egg" paradox of proteins and nucleic acids, which has troubled the molecular biologists for more than half a century. It is seemingly a logical circular debate about which appeared first ("metabolism first" vs "replication first"). On the other hand, any hypothesis cannot evade this paradox.

Amazingly, 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.

Logically, it is impossible that such complex coherent pathways and intertwined processes could have evolved separately. It appears that the primitive biochemical system was developed by molecular cooperation with functional optimization, i.e. coevolution, a term also used for Darwinian selection. The term "symbiosis" is also used to refer to molecular evolution [8].

1-5. The missing "matchmaker"

Here, I put forward a logical reasoning that if "A" can do two things ("B", "C"), then it is possible to establish a connection (precursors of information) between "B" and "C". Meanwhile, other conditions should also be met, for example, such an interaction should take place in an enclosed entity (e.g. a lipid vesicle), and certain selective pressure is also needed to enable a protocell to undergo structural and functional optimizations.

As for the genetic information, "B" and "C" are polynucleotides and polypeptides, respectively (**Figure 1**). In this case, however, "A" is deeply hidden in the biochemical system, and the mark of its past disappeared completely in the evolutionary history of life. Therefore, it is necessary to find the "matchmaker" if we want to know the origin of the genetic code. As mentioned above, macromolecules must have evolved in a closed environment (e.g. lipid vesicle) but never in free solution. If not so, why did nature create a complex cell on Earth?

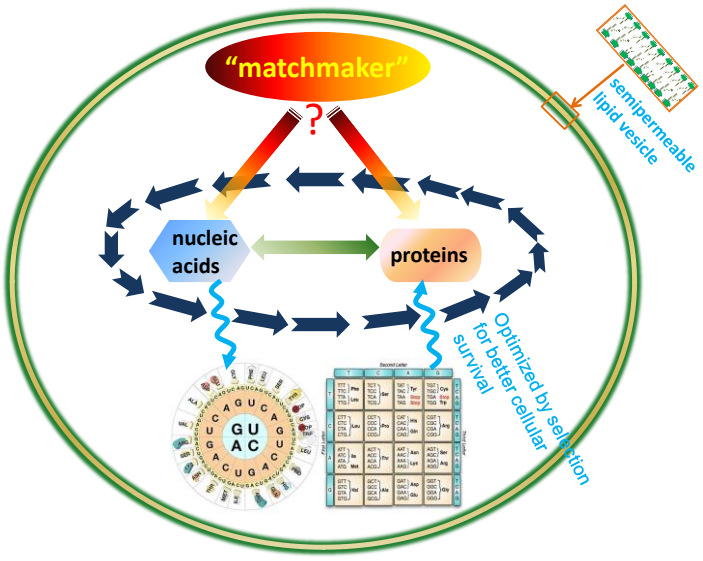

Figure 1. Who is the "matchmaker" between nucleic acids and proteins?

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

It is unfortunate that in all previous theories or hypothesis related to the origin of genetic code, attention has been paid neither to such logical reasoning, nor to the "matchmaker". Consequently, it is impossible to find the right answer only by focusing on the relationship between "B" and "C".

2. 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 2**). ATP is the key to open the black box. ATP was a building block of life universally present in the prebiotic "soup", but currently is renewed by solar-photon driven chemical reactions to prevent exhaustion. In the evolution of the primordial genetic code, ATP would have served as both an energy carrier and an informatization molecule, as it could energetically connect amino acids to short peptides and catalytic polypeptides, and could drive the formation of polynucleotides, resulting 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 it provides energy for the synthesis of all other nucleotides from simple molecules, while A, G, C, and T/U comprise the genetic code.

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 even within a poor primordial biochemistry at the beginning. The genetic code is completed with a triplet codon, perhaps mainly 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. The ATP hypothesis can be viewed as an energetic model for the origin of the genetic code.

The Earth came into being about 4.5 BYA, and the "late heavy bombardment" that lasted until around 4.0 BYA likely precluded life's survival [**22, 23**]. This could be a stage of prebiotic "soup" when no evolution occurred in spite of the presence of relatively complex compounds. With a gradual cooling of the Earth, prebiotic evolution began in a lipid vesicle (precursor of the protocell) where the primordial biochemical system developed, energetically driven by ATP. It is postulated that the first vesicles were formed from conjugated linolenic (C18:3n-3) and parinaric

(C18:4n-3) acids which would form vesicles stable at the high temperatures (~85 ℃) and the somewhat acidic pH values (6.0-6.5) of the Archean ocean surface [24]. The next stage is life evolution (Darwinian evolution) with self-assembling, self-replication and periodic cell division, guided by triplet codon. There is no doubt that in extant biochemical system, ATP drives a set of dynamic processes, e.g., the construction of molecules from smaller parts (anabolism) and the breakdown of the larger ones into smaller parts (catabolism), building up the individual components of the living machine…which is also called as metabolism [25].

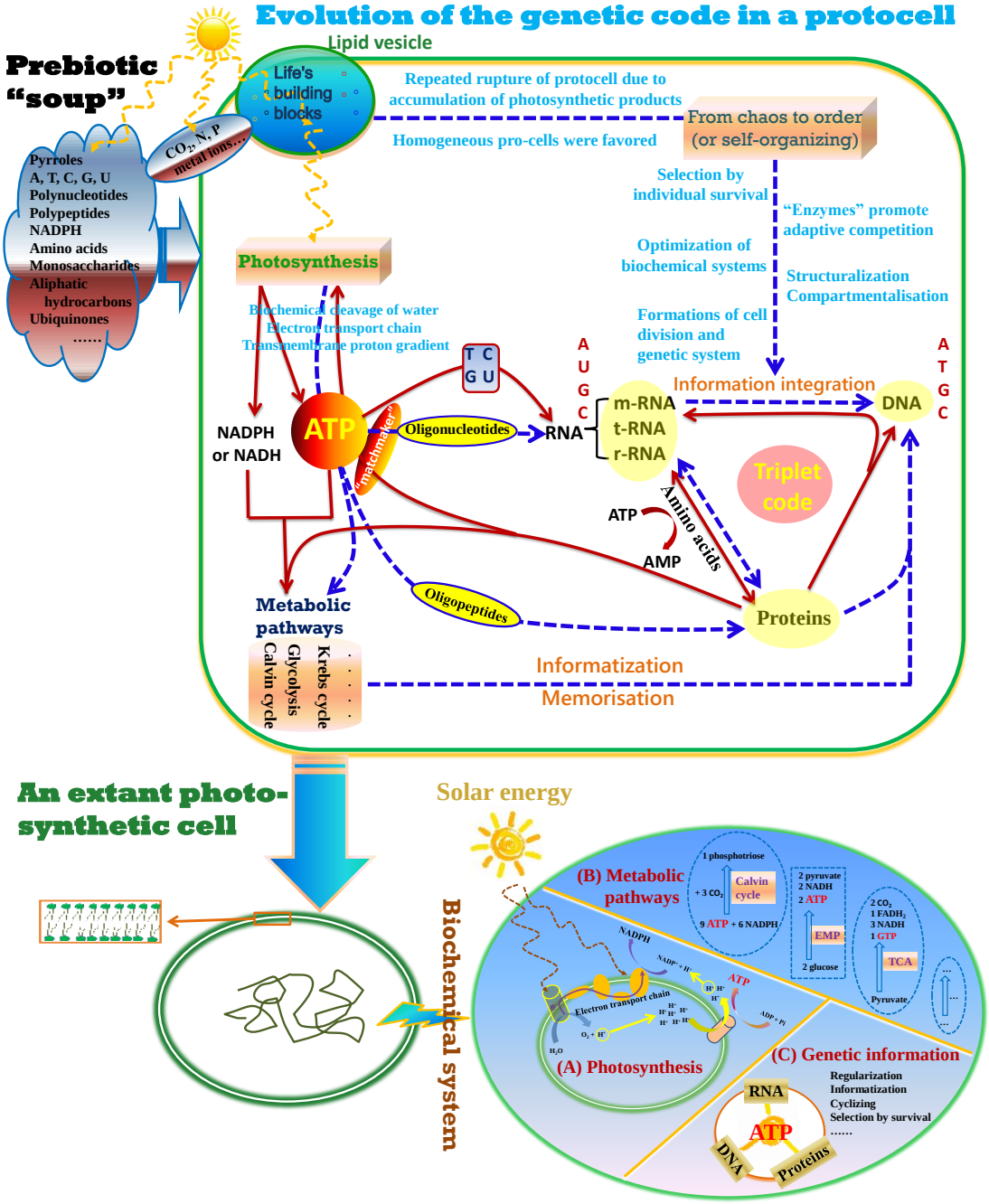

Figure 2. A simplified conceptual model of the origin of the genetic code based on the ATP hypothesis. In the protocell, dashed blue lines indicate evolutionary processes during the precellular period, while solid red lines denote current processes or interactions;

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

and arrows indicate the direction of influences or actions. Energy (e.g. sunlight), building blocks of life and semipermeable lipid vesicles (suitable for cell division) are three indispensable elements for the origin of life. The story is assumed to start 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 informatization, memorisation and structuralization/compartmentalisation. Photochemical production of ATP (including biochemical cleavage of water, electron transport chain, transmembrane proton gradient etc.) was favored by precellular selection, conducive to promoting the origin of the genetic code, a process lasting several hundred million years. Then ATP became the energetic heart through mediating biochemical innovations (e.g. 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 preservation of 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, divisible and replicable, developed an elaborate biochemical system where solar energy was continuously transformed into ATP through photosynthesis and further to support other cellular functions through metabolic pathways, commanded by genetic information in triplet codon.

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).

3. Why ATP?

3-1. "Hidden hand" or natural forces?

We cannot deny that codons and amino acids are linked stereochemically [13]. 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. Importantly, only ATP can make an energetic connection across molecules, organisms, species, and the biosphere.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

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 [26]. 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 development of the genetic code in a primordial cell should be structurally simple and energetically constant. Only after this, peptide/nucleic acid interactions and thermodynamic processes could become more and more complex, and well-organized with optimized structure, paths and functions under selective pressure.

3-2. 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 an extant biochemical system is organized. Consider the following question: what is life? We can say, life is not static, and is a program! We can also say in more detail, life, compositionally, is the unity of matter, energy and information and, dynamically, is the harmonious interplay of material cycling, energy flow and information communication. Among these, 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 clarified that organisms are molecular machines in which inconceivably numerous biochemical reactions take place for the acquisition, conversion and use of energy. Except for chemoautotrophic bacteria, all extant 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 [27]. Such bioenergy is stored in a thioester linkage (as in acetyl-CoA), a phosphate-ester bond to carbon like in acetyl phosphate or a phosphate bond in the adenosine triphosphate (ATP) molecule [28].

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. In this sense, ATP is the only universal currency of biological energy. 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 provides biochemical force to these transformations [29]. In this sense, ATP biochemically governs the information exchange, while 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 [30]. Therefore, as an irreplaceable molecule of both energy transformation and informatization, ATP appears to play energetically a central role in the biochemical system.

3-3. How did life originate on Earth?

The origin of the genetic code is a part of the origin of life. The origins of life and the genetic code are two strongly linked problems, and understanding of one requires understanding of the other.

① An indicator molecule for photochemical origin of life

The importance of ATP in the origin of the genetic code could be attributed to its role in driving 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 [31]. 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 [32].

Let's take a look at the possible evidence from an interesting molecule. It is well known that chlorophyll is a very important molecule for photosynthesis. Interestingly, it seems to be an adduct of 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 [33], 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 [34]. 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 [35]. 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 [31]. 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.

② Varied theories on the origin of life

To date, there have been many hypotheses about the origin of life (**Figure 2**), mostly site- and metabolism-related [32, 36]. There is not yet a general consensus in the origins-of-life community about whether the earliest organisms were

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

photosynthetic, chemosynthetic or heterotrophic.

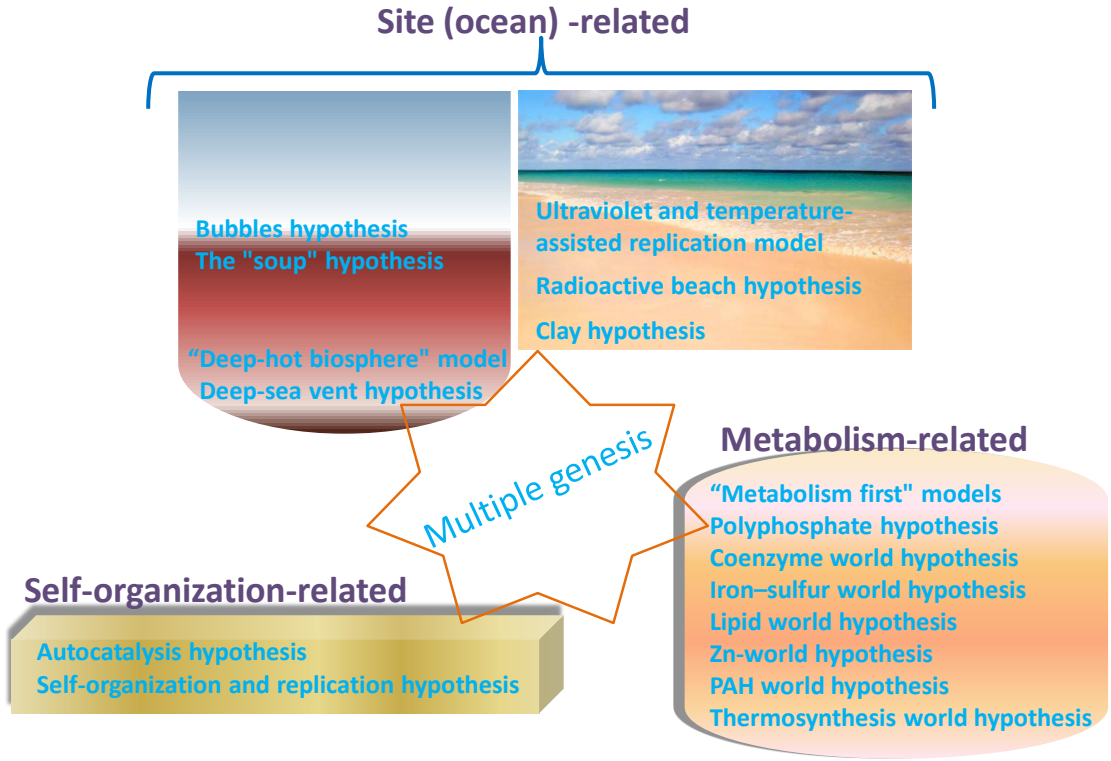

Figure 3. Various hypotheses on the origin of life

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. The source of energy for the earliest life has been strongly disputed, with plausible hypotheses ranging from pH and redox gradients to thermal energy and UV light [36]. It is unfortunate to exclude sunlight, the most universal and constant energy source, probably due to a belief that complex biosystems usually evolved from simpler ones. Yet, simple things are not necessarily older than complex ones, as there is both evolution (complication) and degeneration (simplification). Therefore, the chemosynthetic autotroph was not necessarily the first life.

The ATP hypothesis strongly supports the photochemical origin of life. Studies of the phylogeny of chlorophyll biosynthesis genes are best interpreted as evidence for the ancestral bacterium having been a green bacterium, with the root of the tree of life lying between the sulphur and nonsulphur green bacteria. It is also suggested that the code originated in conjunction with the exploitation of polyphosphate as an energy source by an obcell, whereas its later expansion was stimulated by the evolution of amino acid biosynthesis in a protocell that originated photosynthesis[8]. This is an interesting viewpoint, although I only partially agree with it.

3-4. Concerns on the role of ATP as an initiator of the genetic code

It may be argued that prebiotic origin of pyrimidine nucleotides does not depend on

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

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 [37, 38]. 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 [39], indicating that nucleotides could be present or easily formed in the prebiotic soup. The protocell innovated the photochemical reactions or simple apparatus to produce ATP, which later has been processed efficiently and regularly with the aids of proteins and RNA/DNA, i.e., photosynthesis. This is a result of coevolution, but not a sequential process!

4. How ATP?

4-1. The route from energy transformation to informatization

In the early days of the Earth, there could never be any information, and information was an evolutionary product of cellular life. It is assumed here that the genetic code system was built by a chemical mechanism closely related to the production of ATP in the protocell.

① Transmembrane H^+ gradient for ATP production

Energetically, the first task undertaken by primordial life was likely 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.

Chemically, it was not impossible, on the primitive Earth, that fatty acids could automatically form a double-layered globular membrane structure [40]. 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

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. In this way, membranes began to manage energy through manipulation of the electro-chemical gradient built up between the inside and the outside of the cell (in particular with the fascinating nanomotor ATP synthase) [41].

② Coevolution of macromolecules to facilitate ATP production

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.

③ Non-organized prebiotic "soup" and occasional self-organization

The prebiotic mixtures of chemicals (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 [32]. 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 [21]. It was the first step for the evolution of life to sequestrate various life's 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. In my view, occasional self-organization (e.g. polymerization, self-replication) of molecules should have already existed in the early days of the Earth, although

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

primitive life commenced its journey from materials that were almost not organized.

④ ***How did cellular division begin?***

Accidental encapsulation does not make a cell. First of all, development of cellular division was a prerequisite of the genetic code. Let us first consider the semipermeable lipid vesicle (enclosed by amphiphiles having 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.

It is the capacity of membranes for continuous growth and discontinuous division that makes organisms physically possible. For example, at temperatures where the amphipathic molecules are fluid, the membrane can grow spontaneously by the insertion of additional molecules from the environment and divide spontaneously as a result of natural instabilities or shearing forces, and cells merely harness, direct or modify these basic physical forces [**8**]. Studies on spontaneous evolution of lipid vesicles showed that they split, fuse, get internalised and make complex internal networks [**42**].

⑤ ***Towards informatization***

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 "intrinsic" attributes or function like 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 as informatization) along with cellular compartmentalization (**Figure 2**). 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. This is also a process to innovate the primitive "essence" of life.

4-2. The route from ATP to triplet codon

How did ATP initiate coding between two types of macromolecules in the form of triplet codon? How can we chart the path of information evolution? 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 ***A matchmaker between polynucleotides and polypeptides***

It can be imaged that in the organic "soup" enclosed by the protocell, the energetic ATP with its derivatives could randomly extend chains of both polynucleotides and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

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 macromolecules (i.e., a sequence-specific interaction), that is, information! It is 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 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 and functional differentiation

Informatization was inevitably coupled with structuralization (e.g. structural subdivision or specialization) and functional differentiation, which provided 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 almost all biochemical 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 2**). 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?

It is reasonable to postulate that the triplet codon was just a result of appropriate stereochemical handling of an amino acid and the optimization of biochemical networks under selective pressure. To handle ca. 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

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?***

The individuality of primordial life was determined by the size limitation of a lipid vesicle, optimized by photosynthesis-driven quasi cell division. This resulted in development of slef-organization, division and reproduction of a homogenous protocell, commanded by the genetic code. 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.

It is suggested that only macromolecules that operate on their environment to produce further copies of themselves have an evolutionary future [**43**]. In terms of individuals (protocells), this seems to be a plausible view.

4-3. The route 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 [**18**], 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 [44]. 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.

5. ATP-centered RNA World

Woese [45] proposed the RNA world hypothesis (a term coined by Walter Gilbert in 1986 [46]): 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 [47]. 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 [21, 44, 48].

Laboratory experiments on RNA evolution confirmed the roles of RNA in catalyzing nucleotide synthesis, RNA polymerization, aminoacylation of transfer RNA and peptide bond formation [49, 50, 51, 52]. 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 [53, 54].

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 [55]. 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 [56].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

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 2**). 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 [**57, 58, 59**], 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 [**43**]. Although we still do not know exactly what happened before the emergence of RNA and why, there are numerous evidences on the dependence of RNA on ATP in extant cells, e.g. the synthesis of RNA in isolated thymus nuclei is ATP dependent [**60**]. 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 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.

6. Where To Go?

The origin of life is summarized as “container-enabled chemistry first” or “self-assembling” [**61, 62**]. 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 has 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 [**63**]. 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 maintenance of individuality, as an eco-physiological force, was engaged in reshaping or refixing the rhythm or regularity of biochemical reactions in protocells? Darwinian

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

evolution could be the course that ultimately drove the formation of advanced forms of life. When and how did protocell lines of descent escape from the fate of the dead ends? When and how did a protocell begin to be transferred/copied to its daughter protocells? How many processes were required for a reproductive minimal protocell? These questions are yet unanswered.

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. In my view, 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. Although the present model is rife with assumptions, it is aimed to provide a new conceptual framework, focusing on energetic force (ATP) and coevolution of the genetic code with biochemical system, which is basically different from the major theories (**Table 1**) as well as the other hypotheses (**Supplementary Table 1**). Although coevolution of the genetic code, protein synthesis and nucleicacid replication is discussed, no attention is paid to energetic forces [**64**]. 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 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 stereochemical 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.	[13]
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-evolution	The code was assigned in parallel to the evolution	1. The structure of the codon system is primarily an imprint of the prebiotic pathways of	[2]

theory	of amino acid biosynthesis	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.	
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 chemical interactions between particular amino acids and short nucleic acid sequences, whereas lack of such interactions excluded other amino acids from proteins entirely.	[14]
The ATP theory	The genetic code is a product of ATP-driven evolution of the pristine biochemical system in a lipid vesicle that later developed into a protocell where ATP drove biochemical innovations from energy transformation to informatization, i.e., ATP energetically mediated syntheses of polynucleotides and polypeptides, leading to the emergence of a triplet codon system with cyclization of nucleic acid and proteins into a	1. Energy (sunlight), building blocks of life and semipermeable lipid vesicles (suitable for cell division) are three indispensable elements for the origin of cellular life commanded by the genetic code; 2. ATP is at the biochemical heart of precellular evolution by occupying the position as the only energetic product of photosynthesis translating solar energy to chemical energy, and by driving biochemical innovations (e.g. Krebs cycle, glycolysis, Calvin cycle etc.) to perform intracellular energy transformation; 3. ATP also serves as an informatization molecule, because it could energetically elongated chains of both polynucleotides and polypeptides, thus providing a bridge between these bio-polymers. This drove the biochemical development from energy transformation to informatization, a	The present study

	loop of reciprocal causation.	revolutionary leap in the prebiochemistry; 4. To optimize and memorize biochemical reactions in a primordial cell, ATP-driven informatization co-developed with structuralization, leading to cyclization of polynucleotides and polypeptides into a feedback loop of reciprocal causation, and then the genetic information stored in RNA was permanently fixed to DNA. The triplet codon might be only for stereochemical handling of amino acids through, e.g., Watson–Crick pairing interactions. 5. The first cellular life came into being at least with an ATP-centric biochemical system to transform solar energy to chemical energy, also capable of self-assembling, self-replication and self-division, through a series of biotransformations catalyzed by enzymes.	

Reactions that cannot run spontaneously may occur with the assistance of ATP that has a unique capacity to couple with the formation of most important biopolymers (polypeptides, polynucleotides, polysaccharides, etc.) [65]. In fact, the energetic role of ATP in the origin of life has already been recognized, e.g., ATP was considered as a universal source of energy [66], and a machinery of ordering [65]. It is declared that there must be a certain chemical mechanism producing ordering as a regular, for the given conditions, form of the evolution of matter [65]. I dare say this depended on the ATP-driven evolution of biochemical systems in a primordial cell, undergoing a revolution from energy transformation to informatization (thus triplet codon). It couldn't be anything else!

Crick [67] says, "I hope…that when people put forward detailed theories about the origin of the genetic code, they will try if possible to produce ones which can be tested in some way or other". However, it is challenging to test the credibility of various theories/hypotheses proposed to explain the origin of the genetic code [68]. Fortunately, 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 [69] hypotheses, and so on, reflect each unique aspect of the story. The ATP hypothesis is related, to a greater or lesser extent, with the previous hypotheses (**Figure 4**). Interestingly, the tidal cycling hypothesis suggests that life might emerge from surface ocean, but not from alkaline thermal vents in the sea floor. The ATP hypothesis would greatly improve our understanding of the origin of the genetic code that handles information, a measurable abstract entity [70]. Nevertheless, the collection of supporting evidence to verify this hypothesis remains a challenge.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

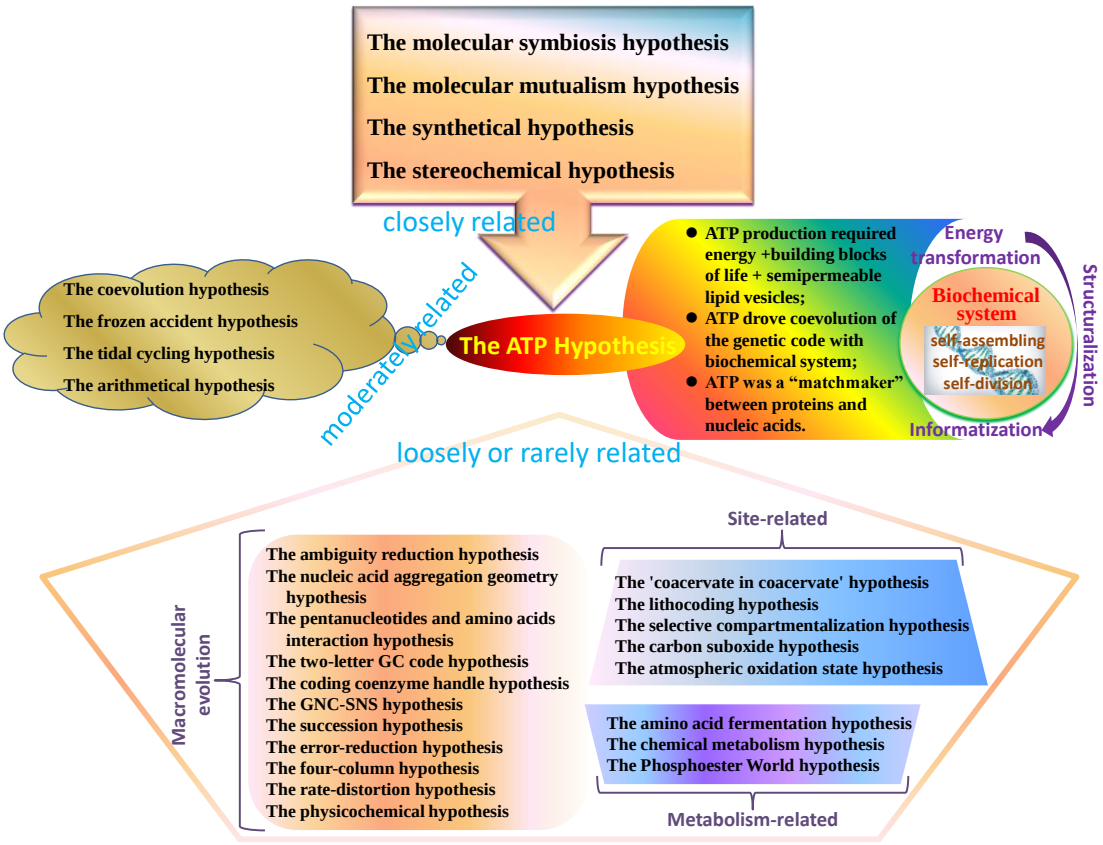

Figure 4. Possible relations between the ATP hypothesis and the other hypotheses on the origin of the genetic code

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, Joseph Moran, Pierre Zofel and Alison Liu for their critical but valuable comments and suggestions. This research was supported by the National Natural Science Foundation of China (91951110) and 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. Gonzalez, D. L. Chapter 6 The mathematical structure of the genetic code. In: Barbieri M. (ed.), The Codes of Life: The Rules of Macroevolution. Springer: Netherlands, 2008.
7. Chatterjee, S.; Yadav, S. The Origin of Prebiotic Information System in the Peptide/RNA World: A Simulation Model of the Evolution of Translation and the Genetic Code. Life **2019**, 9, 25; doi:10.3390/life9010025
8. Cavalier-Smith, T. 2001. Obcells as proto-organisms: membrane heredity, lithophosphorylation, and the origins of the genetic code, the first cells, and photosynthesis. *J. Mol. Evol.* **2001**, 53, 555-595.
9. Rauchfuss, H. Chemical evolution and the origin of life. Springer-Verlag Berlin Heidelberg, 2008.
10. Smith, J. M.; Szathmáry, E. The Major Transitions in Evolution. Oxford University Press: New York, NY, USA, 1995
11. Yockey, H.P. Information theory, evolution, and the origin of life. Cambridge University Press: Cambridge, UK, 2005.
12. Gamow, G. Possible relation between deoxyribonucleic acid and protein structures. *Nature*, **1954**, 173: 318.
13. 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.
14. 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.
15. 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.
16. Koonin, E.V.; Novozhilov, A.S. Origin and evolution of the genetic code: the universal enigma. *Cell Mol. Biol.* **2009**, 61, 99-111.
17. Ohama, T.; Inagaki, Y.; Bessho, Y.; Osawa, S. Evolving genetic code. *Proc. Jpn. Acad. Ser. B.* **2008**, 84, 58–74.
18. 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.
19. Sciarrino, A.; Sorba, P. Codon-anticodon interaction and the genetic code evolution. *Biosystems* **201**3, 111, 175-180.
20. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **1971**, 58, 465-523.
21. Joyce, G.F. The antiquity of RNA-based evolution. *Nature* **2002**, 418. 214-221.
22. Nisbet, E.G.; Sleep, N.H. The habitat and nature of early life. *Nature* **2001**, 409, 1083-1091.
23. Bada, J.L. How life began on earth: a status report. Earth Planet. Sci. Lett. **2004**, 226, 1-15.
24. Michaelian, K., Rodr ́guez, O. Prebiotic fatty acid vesicles through photochemical dissipative structuring. *Rev. Cuba Qu ́Mica* **2019**, 31, 354–370.

25. Danchin, A. From chemical metabolism to life: the origin of the genetic coding process. *Beilstein J. Org. Chem.* **2017**, 13, 1119-1135.
26. Leslie, M. On the origin of photosynthesis. *Science* **2009**, 323, 1286-1287.
27. 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.
28. Sousa, F.L.; Thiergart, T.; Landan, G.; Nelson-Sathi, S.; Pereira, I.A.C.; Allen, J.F.; Lane, N.; Martin, W.F. Early bioenergetic evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2013**, 368, 20130088.
29. Nelson, D.L.; Cox, M.M. Lehninger Principles of Biochemistry (4th ed.). W.H. Freeman, 2004..
30. Battail, G. Information and Life. Springer, 2014.
31. 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.
32. Xie, P. The Aufhebung and Breakthrough of the Theories on the Origin and Evolution of Life. Science Press: Beijing, P.R. China, 2014.
33. Taiz, L.; Zeiger, E. Plant Physiology (4th edition). Sinauer Associates, 2010.
34. Madigan, M.T.; Martinko, J.M.; Stahl, D.A.; Clark, D.P. Brock Biology of Microorganisms (13th ed). Prentice Hall, 2012.
35. Michaelian, K. Thermodynamic dissipation theory for the origin of life. *Earth Syst. Dynam.* **2011**, 2, 37-51.
36. Preiner, M.; Asche, S.; Becker, S. et al. The future of origin of life research: bridging decades-old divisions. *Life* **2020**, 10, 20; doi:10.3390/life10030020
37. Powner MW, Gerland B, Sutherland JD. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature **2009**, 459, 239-42.
38. 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.
39. Wong, Tze Fei JWong JT, Lazcano A. Prebiotic Evolution and Astrobiology. Landes Bioscience, Austin Texas, 2009.
40. Wong, J.T.F.; Lazcano. A. (eds.) Prebiotic Evolution and Astrobiology. CRC Press, 2009.
41. Iino, R.; Noji, H. Operation mechanism of FoF1-adenosine triphosphate synthase revealed by its structure and dynamics. *IUBMB Life* **2013**, 65, 238–246.
42. D'Aguanno, E.; Altamura, E.; Mavelli, F.; Fahr, A.; Stano, P.; Luisi, P. L. Physical routes to primitive cells: an experimental model based on the spontaneous entrapment of enzymes inside micrometer-sized liposomes. Life 2015, 5, 969-996.
43. Orgel, L.E. Evolution of the genetic apparatus: a review. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, 52, 9-16.
44. Doudna, J.A.; Cech, T.R. The chemical repertoire of natural ribozymes. *Nature* **2002**, 418, 222-228.
45. Woese, C.R. The Genetic Code: the Molecular Basis for Genetic Expression. Harper & Row, 1967.
46. Gilbert, W. The RNA world. *Nature* **1986**, 319, 618.
47. 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.
48. 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.

49. Zhang, B.; Cech, T.R. Peptide bond formation by *in vitro* selected ribozymes. *Nature* **1997**, 390, 96-100.
50. Unrau, P.J.; Bartel, D.P. RNA-catalysed nucleotide synthesis. *Nature* **1998**, 395, 260-263.
51. Lee, N.; Bessho, Y.; Wei, K.; Szostak, J. W.; Suga, H. Ribozyme-catalyzed tRNA aminoacylation. *Nature Struct. Biol.* **2000**, 7, 28-33.
52. 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.
53. Lorsch, J.; Szostak, J.W. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **1994**, 371, 31-36.
54. 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.
55. White, H.B. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **1976**, 7, 101-104.
56. 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.
57. 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.
58. Sengupta, S.; Higgs, P.G. Pathways of genetic code evolution in ancient and modern organisms. *J. Mol. Evol.* **2015**, 80, 229–243.
59. Zimmer, C. On the origin of life on earth. *Science* **2009**, 323, 198-199.
60. McEwen, B.S.; Allfrey, V.G; Mirsky, A.E. Dependence of RNA synthesis in isolated thymus nuclei on glycolysis, oxidative carbohydrate catabolism and a type of "oxidative phosphorylation". *Biochim Bioph Acta (BBA)* **1964**, 91, 23-28.
61. Deamer, D. Assembling Life: How can Life Begin on Earth and other Habitable Planets? Oxford University Press: New York, NY, USA, 2019.
62. Bains, W. Getting beyond the toy domain. Meditations on David Deamer's "Assembling Life". Life 2020, 10, 18; https://doi.org/10.3390/life10020018.
63. Mayer, C. Life in the context of order and complexity. *Life* **2020**, 10, 5, https://doi.org/10.3390/life10010005.
64. Alberti S. Evolution of the genetic code, protein synthesis and nucleic acid replication. *Cell. Mol. Life Sci.* **1999**, 56, 85-93.
65. Galimov, E. M. Phenomenon of life: between equilibrium and non-linearity. Origin and principles of evolution. *Geochem. Int.* **2006**, 44(Suppl. 1), S1-S59.
66. de Duve, C. Life Evolving - Molecules, Mind, and Meaning. Oxford University Press: New York, NY, USA, 2002.
67. Crick, F.H.C. The Genetic Code—Yesterday, Today, and Tomorrow. *Cold Spring Harb. Symp. Quant. Biol.* **1966**, 31, 3-9.
68. 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.
69. Barbieri, M. Evolution of the genetic code: the ambiguity-reduction theory. *BioSystems* **2019**,

185, 104024.

70. Battail, G. Error-correcting codes and information in biology. *BioSystems* **2019**, 184, 103987.

Supplementary Table 1 A list of the other hypotheses on the origin of genetic codes

Name	Connotation	Ref.
The physicochemical hypothesis	The driving force behind the origin of the genetic code structure was the one that tended to reduce the physicochemical distances between amino acids codified by codons differing in one position.	[13, 71]
The ambiguity reduction hypothesis	Groups of related codons were assigned to groups of structurally similar amino acids and the genetic code, therefore, reached its current structuring through the reduction of the ambiguity in the coding between and within groups of amino acids.	[72,73, 74]
The selective compartment-alization hypothesis	An oil-slick covered the surface of the primitive ocean, constituents of which formed association colloids or micelles at the water-oil-air interfaces. Depending on the polarity of the media, these aggregates possessed hydrophilic and hydrophobic interiors where selective uptake of amino acids and nucleic acid constituents could take place. Condensation and polymerization in the micellar phase were enhanced, eventually leading to the formation of a charged adaptor loop with an anticodon which is complementary to the presently valid codon. Only two groups of amino acids, hydrophilic and hydrophobic, were recognized by the primitive translation mechanism.	[75]
The 'coacervate in coacervate' hypothesis	The sequence of the bases in the primordial nucleic acid molecules was the outcome of regular and progressive evolution at a given level of chemogenesis that occurred every- where the necessary chemicals and environmental conditions prevail. Even if the primary abiogenetic proteins originated without a nucleic acid matrix, they presumably had a regular, non-fortuitous base, although they were no doubt far less constant than recent, biogenic proteins. Primary nucleic acid molecules undoubtedly originated where these proteins accumulated in the presence of further necessary materials and other conducive conditions. It was these hereditary 'coacervates in coacervate' which underwent further evolution through successive mutations of the genetic code of their DNA molecules.	[76]
The carbon suboxide hypothesis	Carbon suboxide (C_3O_2) polymers formed in the primitive atmosphere would have produced an organic compound 'soup' of high concentration on the Earth. Various vestiges of C_3O_2 are found in the present genetic scheme, which might suggest that the living system had formed from the polymer 'soup'.	[77]
The nucleic acid aggregation geometry hypothesis	The primitive translational complex responsible for the evolutionary origin of the genetic code was an ordered RNA aggregate, stabilized not only by base-pairing, but also by the multivalent ionic interactions known to be responsible for bundle aggregation and condensation of DNA. A triplet code is the logical consequence of the structural	[78]

	symmetry and information coding capacity and efficiency of the proposed class of aggregates.	
The pentanucleotides and amino acids interaction hypothesis	A particular conformation of a pentanucleotide forms a double sided template, with its 'inside' capable of nestling an amino acid while the 'outside' acts as an adaptor to a 'codon' triplet on long-chain nucleic acids. This serves as a primitive decoding system. A dynamic interaction is triggered, by this decoding system, through which amino acids are brought to juxtaposition facilitating peptide bond formation.	[79]
The two-letter GC code hypothesis	The initial genetic code consisted of only two letters, G and C, and then expanded the number of available codons via the introduction of an additional pair of letters, A and U.	[80, 81]
The coding coenzyme handle hypothesis	Useful coding preceded translation. Early adapters, the ancestors of present-day anticodons, were charged with amino acids acting as coenzymes of ribozymes in a metabolically complex RNA world. The ancestral aminoacyl-adapter synthetases could have been similar to present-day self-splicing tRNA introns. A codon-anticodon-discriminator base complex embedded in these synthetases could have played an important role in amino acid recognition. Extension of the genetic code proceeded through the take-over of nonsense codons by novel amino acids, related to already coded ones either through precursor-product relationship or physicochemical similarity.	[82]
The GNC-SNS hypothesis	The genetic code originated from the GNC code by acquiring the four amino acids, [G, A, D, V], which could be easily synthesized and accumulated in large quantities by the prebiotic systems on primitive earth. After that, the GNC code evolved to the GNS and SNS codes, successively. Finally, the universal genetic code was formed by successfully capturing A and U-initiated codons.	[83]
The succession hypothesis	Three successive levels of chemical specificity generated the nucleotide assignments of amino acids in the genetic code. The first level results from hydrophobic and stereospecific interactions between amino acids and short oligonucleotides (termed oligons). The second and third levels of specificity are determined by conditions of energy transfer from loaded oligons (amino acid–oligomer covalently linked) to formation of phosphodiester bond (second level of specificity) and peptidic bond (third level of specificity), while these reactions are catalyzed by RNA templates. Simple physical processes, in which a level of specificity is integrated in an emerging meta-structure expressing new properties, generate a parsimonious and realistic explanation of emergence of the genetic code.	[84]
The atmospheric oxidation state hypothesis	At the period I (preoxygen), in the primary Earth atmosphere the first nitrogen base, adenine (A), containing no oxygen appeared. The period II (oxygenated), during which three other nitrogen bases appeared in the atmosphere, consisted of three stages; at the first stage, guanine (G)	[85]

	appeared, at the second, cytosine (C), and at the third stage, uracyl (U). In accordance with the above periods, formation of codons and amino acids in nature was taking place presumably by the following way: at the period I, the first and the only codon AAA appeared, to which the amino acid lysine (Lys) corresponded; at the first stage of the period II, 7 codons and 3 amino acids (Arg, Glu, Gly) appeared; at the second stage, 19 codons and 8 new amino acids (Asn, Gin, Ser, Asp, Thr, Ala, His, Pro) appeared; at the third stage, 37 codons and more 8 new amino acids (Trp, Tyr, Cys, Ile, Met, Val, Leu, Phe) appeared. Thereby, in the course of biochemical evolution, 20 amino acids and 64 codons appeared in nature.	
The tidal cycling hypothesis	Tide-driven cycles of concentration and dilution generated, and then amplified, primitive nucleic acid polymers, and also led to the emergence of template-directed polypeptide assembly. Such cycling could also underlie the development of cell-based life. Fast tidal cycling could establish a TCR (tidal chain reaction) analogous to a PCR (polymerase chain reaction) acting on nucleic acid polymers, allowing their self-replication.	[86, 87, 88]
The error-reduction hypothesis	The amount of tolerance that the evolving assignment of codons to amino acids developed for a particular type of genetic error was in proportion to the prevalence of this type of error. The relative rates of the various types of genetic errors should have left characteristic imprints in the structure of the genetic code. For thermodynamic reasons, RNA in thermophiles tends to possess elevated G+C content. As the frequencies of G and C in the genome were not elevated, the genetic code was fixed in organisms that were not thermophilic or not in an RNA world.	[89]
The arithmetical hypothesis	The genetic code turns out to be a syntactic structure of arithmetic, the result of unique summations that have been carried out by some primordial abacus at least three and half billion years ago. The decimal place-value numerical system with a zero conception was used for that arithmetic. It turned out that the zero sign governed the genetic code not only as an integral part of the decimal system, but also directly as an acting arithmetical symbol. Being non-material abstractions, all the zero, decimal syntax and unique summations can display an artificial nature of the genetic code. They refute traditional ideas about the stochastic origin of the genetic code. A new order in the genetic code hardly ever went through chemical evolution and, seemingly, originally appeared as pure information like arithmetic itself.	[90]
The four-column hypothesis	The earliest amino acids in the code were those that are easiest to synthesize non-biologically, namely Gly, Ala, Asp, Glu and Val. These amino acids are assigned to codons with G at first position. Therefore the first code may have used only these codons. The code rapidly developed into a four-column code where all codons in the same	[91]

	column coded for the same amino acid: NUN = Val, NCN = Ala, NAN = Asp and/or Glu, and NGN = Gly. Later amino acids were added sequentially to the code by a process of subdivision of codon blocks in which a subset of the codons assigned to an early amino acid were reassigned to a later amino acid. Later amino acids were added into positions formerly occupied by amino acids with similar properties because this can occur with minimal disruption to the proteins already encoded by the earlier code. As a result, the properties of the amino acids in the final code retain a four-column pattern that is a relic of the earliest stages of code evolution. The driving force during this process is not the minimization of translational error, but positive selection for the increased diversity and functionality of the proteins that can be made with a larger amino acid alphabet.	
The molecular mutualism hypothesis	Synthesis of complex organic monomers and polymerization reactions occurred within a surface hydrophilic layer and at its aqueous and atmospheric interfaces. Replication of nucleic acids and translation of peptides began at the emulsified interface between hydrophobic and aqueous layers. At the core of the protobiont was a family of short nucleic acids bearing arginine's codon and anticodon that added this amino acid to pre-formed peptides. In turn, the survival and replication of nucleic acid was aided by the peptides. The arginine-enriched peptides served to sequester and transfer phosphate bond energy and acted as cohesive agents, aggregating nucleic acids and keeping them at the interface.	[92]
The rate-distortion hypothesis	The genetic code originated as a result of the interplay of the three conflicting evolutionary forces: the needs for diverse amino-acids, for error-tolerance and for minimal cost of resources. Stickland fermentation occurred in the RNA world, and pairs of complementary proto-adapters were assigned to Stickland amino acids pairs.	[93]
The amino acid fermentation hypothesis	There is evidence that the genetic code was established prior to the existence of proteins, when metabolism was powered by ribozymes. Also, early proto-organisms had to rely on simple anaerobic bioenergetic processes. Amino acid fermentation powered metabolism in the RNA world, and this was facilitated by proto-adapters, the precursors of the tRNAs.	[94]
The lithocoding hypothesis	This hypothesis assumed the direct contact of pairs of coding molecules with amino acid side chains in hollow unit cells (cellules) of a regular crystal-structure mineral. The coding nucleobase-containing molecules in each cellule (named ''lithocodon'') partially shield each other; the remaining free space determines the stereochemical character of the filling side chain. Apatite-group minerals are considered as the most preferable for this type of coding (named ''lithocoding''). The magmatic nature of the mineral, abiogenic synthesis of organic molecules and polymerization events are considered within the	[95]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202003.0419.v8

	framework of the proposed "volcanic scenario".	
The chemical metabolism hypothesis	Trying to understand the origins of life should be based on what we know of current chemistry in the solar system and beyond. There, amino acids and very small compounds such as carbon dioxide, dihydrogen or dinitrogen and their immediate derivatives are ubiquitous. Surface-based chemical metabolism using these basic chemicals is the most likely beginning in which amino acids, coenzymes and phosphate-based small carbon molecules were built up. Nucleotides, and of course RNAs, must have come to being much later.	[25]
The Phosphoester World hypothesis	The early metabolism arising in a Thioester world gave rise to amino acids and their simple peptides. The catalytic activity of these early simple peptides became instrumental in the transition from Thioester World to a Phosphate World. This transition involved the appearances of sugar phosphates, nucleotides, and polynucleotides. The coupling of the amino acids and peptides to nucleotides and polynucleotides is the origin for the genetic code.	[96]
The molecular symbiosis hypothesis	In the peptide/RNA world, lipid membranes randomly encapsulated amino acids, RNA, and peptide molecules, which are drawn from the prebiotic soup, to initiate a molecular symbiosis inside the protocells. This endosymbiosis led to the hierarchical emergence of several requisite components of the translation machine: transfer RNAs (tRNAs), aminoacyl-tRNA synthetase (aaRS), messenger RNAs (mRNAs), ribosomes, and various enzymes. When assembled in the right order, the translation machine created proteins, a process that transferred information from mRNAs to assemble amino acids into polypeptide chains. This was the beginning of the prebiotic information age. The genetic code developed in three stages that are coincident with the refinement of the translation machines: the GNC code that was developed by the pre-tRNA/pre-aaRS/pre-mRNA machine, SNS code by the tRNA/aaRS/mRNA machine, and finally the universal genetic code by the tRNA/aaRS/mRNA/ribosome machine.	[7, 8]

Supplementary references

71. Sonneborn, T.M. Degeneracy of the genetic code, extant, nature, and genetic implications. In: Bryson, V.; Vogel, H.J. (eds.) Evolving Genes and Proteins. Academic Press, New York, pp. 377-397, 1965.
72. Woese, C. R. On the origin of the genetic code. *Proc. Natn. Acad. Sci. U.S.A.* **1965**, 54, 1546–1552.
73. Fitch, W. M. The relation between frequencies of amino acids and ordered trinucleotides. *J. Mol. Biol.* **1966**, 16, 1-8.
74. Fitch, W. M., Upper, K. The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, 52, 759–767.

75. Nagyvary, J; Fendler, J. Origin of the genetic code: a physical-chemical model of primitive codon assignments. *Orig. Life* **1974**, 5, 357-362.
76. Novak, V.; Liebl, V. On the question of the origin and evolution of the genetic system. *Orig. Life* **1975**, 6, 269-271.
77. Shimizu, M. Carbon suboxide and the genetic code. *Astroph. Space Sci.* **1979**, 62, 509-513.
78. Crothers, D.M. Nucleic acid aggregation geometry and the possible evolutionary origin of ribosomes and the genetic code. *J. Mol. Biol.* **1982**, 162, 379-391.
79. Balasubramanian, R. Origin of life: A hypothesis for the origin of adaptor-mediated ordered synthesis of proteins and an explanation for the choice of terminating codons in the genetic code. *Biosystems*, **1982**, 15, 99-104.
80. Hartman, H. Speculation of the evolution of the genetic cod III: the evolution of t-RNA. *Orig. Life* **1984**, 14, 643 648.
81. Frank, A., Froese, 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
82. Szathmáry, E. Coding coenzyme handles: a hypothesis for the origin of the genetic code. *Proc. Natl. Acad. Sci. USA* **1993**, 90, 9916-9920.
83. Ikehara, K.; Omori, Y.; Arai, R.; Hirose, A. A novel theory on the origin of the genetic code: a GNC-SNS hypothesis. *J. Mol. Evol.* **2002**, 54, 530–538.
84. Seligmann, H; Amzallag, G.N. Chemical interactions between amino acid and RNA: multiplicity of the levels of specificity explains origin of the genetic code. *Naturwissenschaften* **2002**, 89, 542–551.
85. Shabalkin, I.P.; Shabalkin, P.I.; Yagubov, A.S. Evolution of genetic alphabet and of amino acid code. *J. Evol. Biochem. Physiol.* **2003**, 39, 608-615.
86. Lathe, R. Fast tidal cycling and the origin of life. *Icarus* **2004**, 168:18-22.
87. Lathe, R. Tidal chain reaction and the origin of replicating biopolymers. *Int. J. Astrobiol.* **2005**, 4, 19-31.
88. Lathe, R. Tidal cycling and the origin of the genetic code: implications for cellular Life. J. Seckbach (ed.), Genesis - In The Beginning: Precursors of Life, Chemical Models and Early Biological Evolution. Cellular Origin, Life in Extreme Habitats and Astrobiology **2012**, 22, 691-707.
89. Gutfraind, A.; Kempf, A. Error-reducing structure of the genetic code indicates code origin in non-thermophile organisms. *Orig. Life Evol. Biosph.* **2008**, 38, 75-85.
90. shCherbak, V. Chapter 7 The arithmetical origin of the genetic code. In: Barbieri M. (ed.), The Codes of Life: The Rules of Macroevolution. Springer, 2008.
91. Higgs, P.G. A four-column theory for the origin of the genetic code: tracing the evolutionary pathways that gave rise to an optimized code. *Biol. Direct* **2009**, 4:16.
92. 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.
93. Tlusty, T. A colorful origin for the genetic code: Information theory, statistical mechanics and the emergence of molecular codes. *Phys. Life Rev.* **2010**, 7, 362-376.
94. de Vladar, H.P. Amino acid fermentation at the origin of the genetic code. *Biol. Direct* **2012**, 7:6.
95. Skoblikow, N.E.; Zimin, A.A. Hypothesis of lithocoding: origin of the genetic code as a ‘‘Double Jigsaw Puzzle’’ of nucleobase-containing molecules and amino acids assembled by

 doi:10.20944/preprints202003.0419.v8

sequential filling of apatite mineral cellules. *J. Mol. Evol.* **2016**, 82, 163–172.

96. Hartman, H.; Smith, T.F. Origin of the genetic code is found at the transition between a Thioester World of Peptides and the Phosphoester World of Polynucleotides. *Life*, **2019**, 9, 69; doi:10.3390/life9030069.