The Protein/RNA World and the Origin of Life

Sankar Chatterjee

Department of Geosciences, Museum of Texas Tech University, Box 4319, 3301 4th Street, Lubbock, Texas 79409, USA
* Correspondence: Sankar.chatterjee@ttu.edu; Tel: +1-806-787-4332

Abstract: The transition from the *Peptide/RNA world* to the *Protein/RNA world* in the hydrothermal vent environment was a major event in the history of life. The advent of proteins utterly changed the conditions of emerging life, representing a watershed in its development. During subsequent translation various protein enzymes emerged driving protocells into a more complex and interconnected system. With their astonishing versatility, the protein enzymes catalyzed crucial biochemical reactions within protocells into more complex biomolecules in diverse metabolic pathways, whereas structural proteins provided strength and permeability in the cell membrane. Four major events followed after availability of various kinds of protein molecules during prebiotic synthesis. These are: (1) the modification of the phospholipid membrane into the plasma membrane; (2) the origin of primitive cytoplasm; (3) the beginnings of the virus world; and (4) the advent of DNA. The first innovation mediated by proteins was the improvement of the cell membrane. The phospholipid membrane was initially evolved in a vent environment from the gradual modification of a fatty acid membrane via an intermediate phosphatidate acid by non-enzymatic reactions. The phospholipid is then synthesized from phosphatidate acid by a series of enzymes. To make the phospholipid membrane more permeable, various protein molecules interacted with the cell membrane. Proteins not only stabilized the wall membrane, but also acted as pumps, preventing some molecules from the protocells from crossing the membrane barriers, while permitting other selected molecules and ions to enter and leave the protocell. The second modification led by proteins is the gradual conversion of the interior of the protocell from a water-like medium into a gel-like cytoplasm, which became the storehouse of a wide range of biomolecules including amino acids, proteins, nucleic acids, ribosomes, as well as salt and water. The third innovation utilizing the newly synthesized proteins was the emergence of the ancient virus world. In the milieu of different kinds of mRNAs in the prebiotic soup, jelly-roll capsid genes originated *de novo* within genomes of nonviral mRNAs by overprinting. These fragile capsid genes were possibly coated by proteins on the mineral substrate for stability and durability, transforming them into ancient viral particles. These protein coats were random and were not encoded by encased genes. Some protocells might have engulfed these viral particles, when the capsid genes utilized the ribosomes of the host to translate into the appropriate capsid proteins. These capsid proteins then coated the viral genes to make new copies of primordial viruses inside the protocell. Since then, viruses became capsid-encoding organisms. These primordial mRNA viruses parasitized RNA-based protocells, manipulating them to make new copies of themselves. This was the beginning of a relentless war between viruses and their protocellular hosts. The next stage in viral evolution was the emergence of a primitive retrovirus (pre-retrovirus) with a new kind of replicative strategy in a sense that it could turn its RNA into DNA using its own reverse transcriptase enzyme. This is the

beginning of the *Retro world* that facilitated the transition from RNA to DNA genomes. The infection of RNA protocells with pre-retroviruses progressively transferred the RNA genome to a viral DNA genome by retro-transcription. The advent of DNA by the pre-retrovirus marks the fourth innovation, when a number of enzymes had already developed and were utilized by pre-retroviruses. With continued infection, DNA viruses slowly transferred not only their core replication enzymes, such as helicase, primase, and DNA polymerase, to RNA protocells, but also to their DNAs as well. Thus, began the *DNA world*, when DNA replaced RNA as the major genome of the protocells. With the advent of DNA, replication of information was entirely dissociated from its expression. Because DNA is much more stable than mRNA with more storage capacity, it is a superb archive for information systems in the form of base sequences. DNA progressively took over the replicative storage function of mRNA, leaving the latter for protein synthesis. The new protocell with the DNA genome will diversify into large populations of DNA protocells that will outcompete populations of RNA protocells. Genetic information began to flow from DNA to mRNA to protein in a two-step process involving transcription and translation. In the *biological* stage, DNA replication was central to the binary fission of the first cell, orchestrated by the duplication of genomes and then the division of the parent cell into two identical daughter cells. It was carried out by a set of enzymes that formed a Z-ring at the site of replication. With the onset of binary fission, the population of primitive cells grew rapidly in the hydrothermal vent environment, undergoing Darwinian evolution and diversification. These primordial hyperthermophiles, presumably the first life, obtained food and energy directly from the vent environment. However, such a situation was self-limiting, so the early cells evolved their own mechanisms for generating metabolic energy and synthesizing the molecules necessary for their reproduction. The earliest fossil record ($\geq$ 3.5 Ga) of biotic activity is preserved in the Archean hydrothermal and sedimentary rocks of the Nuvvuagittuq Craton of Canada, the Isua Craton of Greenland, the Pilbara Craton of Australia, the Kaapvaal Craton of South Africa, and the Singhbhum Craton of India in the form of the carbonaceous remains of microbial cells, cellular microfossils, and stromatolites. These microscopic fossils provide crucial evidence of the origin and early evolution of prokaryotic cells, beginning with hyperthermophiles. Molecular phylogenetic analysis suggests that both domains of life – Bacteria and Archaea probably split from the last universal common ancestor (LUCA), a hyperthermophilic organism. In the younger sequences of these Archean cratons, two kinds of photosynthetic bacteria, anoxygenic green sulfur bacteria and oxygenic cyanobacteria, appeared in quick succession from the thermophilic ancestor, indicating a shift of niche from a benthic to a planktonic, with reduced thermotolerance. The development of anoxygenic and oxygenic photosynthesis would have allowed life to escape the hydrothermal setting and invade a newly evolved habitat—broad continental shelves to tap solar energy. Cyanobacteria invaded the global ocean, turned it into blue and green, produced oxygen for the first time, and left their signatures in the carbonates and stromatolites.

Keywords: protein/RNA world; plasma membrane; cytoplasm; virus world; pre-retro virus; emergence of DNA; transcription and replication; first cells; hyperthermophiles; LUCA; bacteria and archaea; anoxygenic bacteria; oxygenic bacteria; global distribution of cyanobacteria

1. Prelude to the Origin of Life

Two perennial questions have long fascinated human beings for millennia: does life exist elsewhere in the universe, and how did life begin on Earth? These are the central questions of the new interdisciplinary science of Astrobiology. Our planet supports life because its terrain is rocky, and it is not too hot or too cold, just the right distance from the Sun to allow water to exist on the surface in liquid form. The discovery that microbial life is abundant and diverse in extreme environments is one of the most important discoveries of recent times. The lifestyle of extremophiles in harsh environments on Earth, exposed to heat, cold, pressure, darkness, toxic water, and/or salty niches may offer the most important clues to the possibility of extraterrestrial life in the inhospitable habitats of Mars, Jupiter's moon Europa, and Saturn's moons Titan and Enceladus. Much of the Universe is clearly hostile to life, and only rare places offer even potential oases for its existence.

Life is a planetary phenomenon. There is not yet any direct evidence of life on other planets in our Solar System or galaxy, but recent NASA explorations have prompted optimism that habitable worlds exist throughout the Milky Way. Is Earth the only planet that harbors life in a Universe of at least 150 billion galaxies and a galaxy with about 200 billion stars? Probably not. There is a strong possibility that life also emerged beyond Earth whenever the necessary physical and chemical conditions were met. Recently, the Kepler space telescope has discovered about 4,000 Earth-like exoplanets in the galaxy, each orbiting a Sun-like star in the 'Goldilocks zone'—that is, at just the right distance for liquid water and life. Some of these exoplanets may be habitable and may hopefully nurture life. It is quite possible that life is a natural consequence of planetary evolution and a 'cosmic imperative' anywhere that habitable zones of liquid water and the right mix of organic molecules are available for even short periods of geologic time [1].

The Universe may be awash with life, but interstellar distances are so staggering that we may never get confirmation of this, unless extraterrestrial intelligent life somewhere begins to send out messages that we can receive and decode. The question is no longer, is there life beyond Earth? The question is, how do we find biosignatures anywhere in the Solar System and beyond through remote sensing?

Abiogenesis, or the origin of life, is the natural process by which life has arisen from non-living matter, such as simple organic compounds. While the details of this process are still shrouded in mystery, the prevailing scientific hypothesis is that transition from non-living to living entities was not a single event, but a gradual process of increasing complexity. Miller-type experiments in the 1950s simulating the early Earth conditions have shown that complex organic molecules can be produced by physical process from simply reducing gases available in the primordial atmosphere. Many variations of this experiment followed. After 60 years of experiments in numerous laboratories around the world, few breakthroughs followed Miller's ground-breaking experiment, and frustrations mounted regarding the production of life in a 'test tube.' Perhaps the early Earth environment simulated in the Miller-type experiment was not the accurate model. The early atmosphere of Eoarchean Earth was dominated by a mixture of carbon dioxide and nitrogen and its environment was more violent from the constant threat of meteoritic impacts, not tranquil as assumed in the Miller experiment [2,3]. If the essential building blocks of life were not synthesized on young Earth, they might have come from outer space, delivered by meteorites. In recent times, with the exploration of space, the study of the origin of life has shifted from Miller-type experiments to a broader perspective including planetary beginnings and astrobiology.

Recent progress in planetary science suggests that our planet at the time of life's origin was a world of global oceans punctuated by small continental islands resembling Seychelles, Madagascar, Great Britain, New Zealand, and Sri Lanka [4,5]. To understand how life could begin on young Earth, it is essential to know what organic compounds–the essential building blocks of life–were likely to have been available and how they interacted in a suitable environment. Using the early history of the Solar System as the starting point, life probably arose in Archean Earth through five hierarchical stages of increasing molecular complexity about 4 billion years ago—*cosmic*, *geologic*, *chemical*, *information*, and *biological* (Fig. 1) [4,6,7]. Each level in the hierarchy represents an increase in organizational complexity, each stage merged smoothly and continuously with the following stage. Each major stage appears to have an evolution of mechanisms that control selection at the preceding, lower level organization. Reversion of the lower level stage can be deleterious to the higher level, complex entities. This multilayered hierarchy provided a kind of quality control at each assembly stage. The clues for these hierarchies come from outer space, from the early Earth, from the neighboring Moon and inner planets, from laboratory experiments, and, especially from cellular components of life itself.

<Figure 1 about here>

All living organisms, from microbes to humans, are made of the same basic building blocks that are assembled into four different types of carbon-based organic molecules. These four types of polymers that make up the majority of cell structures are carbohydrates, lipids, nucleic acids and proteins. Carbohydrates or sugars exist in cells both as single sugar molecules, such as glucose, and as complex molecules, such as starch and cellulose. Sugars are an important source of energy. They form the backbone with phosphorous in nucleic acids. They also provide structural support for cells with communication between cells. Lipids are a highly variable group

of molecules that form fatty acids. The most important role of lipids is as the main component of cell membranes. There are two types of nucleic acids that are essential to life as information molecules. These are RNA and DNA. The final four molecules of life are proteins, which are made from linking amino acids. Proteins are large complex molecules that play many critical roles in cells. They perform a vast array of functions, including catalyzing metabolic reactions, DNA replication, providing structure to the cells, and transporting molecules from one location to another. Here we try to understand the origin of these molecules of life in young Earth.

Life, a local phenomenon of the Earth's surface, can in fact be understood only in its cosmic connection. It formed itself out of stardust, shortly after Earth formed. It is difficult to develop a full understanding of life on Earth without understanding its links to its cosmic environment. In the *cosmic* stage, the building blocks of life could have their beginnings in the tiny icy grains that make up the gas and dust in interstellar space. These organic molecules were formed as a byproduct of a nearby supernova explosion, many times more massive than our Sun. The supernova expelled this material across space to form interstellar clouds, which are factories of organic molecular synthesis. During the formation of the Solar System, this interstellar organic material was integrated into interstellar dusts, comets and carbonaceous asteroids. All terrestrial planets have been seeded with prebiotically significant organic compounds in the early history of our Solar System through the impact of meteorites. The large quantities of organic molecules delivered to the newly formed crust of young Earth during the heavy bombardment phase played a key role to life's origin [8-11].

The Stardust spacecraft, NASA's comet sampling mission, detected these building blocks of life in interstellar dust particles [12]. The presence of sugars and their derivatives in carbonaceous meteorites together with amino acids, nucleobases, amphiphiles and other

compounds of biological importance may have contributed to the inventory of organics that played a role in the emergence of life on Earth. The 5-carbon monosaccharide ribose is an important component of coenzymes (e.g., ATP, FAD, and NAD) and the backbone of the genetic molecule RNA. The related deoxyribose is a component of DNA. The origin of liquid water in young Earth, which is critical to life, is from carbonaceous chondrites. The chemical building blocks arose in an unusual, extreme environment at cold temperature in an interstellar medium with very low density and zero gravity, which was not present in early Earth. Earth was extremely deficient in such organic compounds in its early history. This is why the Miller-type experiments failed to duplicate the synthesis of the building blocks of life in the laboratory because these experiments simulated the environment of the primordial Earth rather than the interstellar environment.

The Murchison meteorite, a carbonaceous chondrite that fell in Australia in 1969, has yielded many molecules of life including sugar, phosphate, nucleobases, amino acids and membrane-forming compounds [10,13]. Similarly, comets contain prodigious quantities of building blocks [14]. It is likely that substantial amounts of such organic compounds were delivered to young Earth from space during meteorite bombardment [16]. In this view, the origin of life is an extraordinary event, the product of two worlds: the exogenous delivery of building blocks of life by meteorites to early Earth, and endogenous production of the first life in the cradles of our planet [9]. The origin of life may have interstellar beginnings during planetary formations, but meteorite impacts in young Earth may have delivered these biomolecules and jump-started life [4,7]. To paraphrase Carl Sagan, life is made of stardust—from the debris of supernova explosion.

The accumulation and concentration of cosmic organic molecules on a planetary surface was an essential early step for the origin of life. But where did life begin on early Earth? The answer may be found in the *geologic stage*. The Hadean history (~4.6 to 4 Ga) is barely recorded in the geologic record because asteroids were continuously battering the early Earth. The heat generated from those intense impacts led to the large-scale melting of the Earth surface, forming a vast molten globe that obliterated most traces of early history [3]. During the beginning of the Eoarchean, as the frequency of impacts slowed down, the molten Earth cooled, clouds formed, and the crust hardened, blanketing the entire planet with dark basaltic rocks. The geologic record indicates that Earth was a water-covered planet about four billion years ago, or earlier, with scattered sialic protocontinents [17]. The lands were barren and the waters lifeless, but the scene was far from calm. It was the Late Heavy Bombardment on young Earth (4.1 to 3.9 Ga) that is most likely the driving force for the origin of life [7] During the intense meteoritic impacts, carbonaceous chondrites delivered both the building blocks of life and water to the planetary surface creating innumerable crater basins [4,18,19]. Large impact basins on the Moon, Mercury, and Mars provide strong evidence for this period of bombardment. Unlike our planetary neighbors, plate tectonics and erosion have completely erased this early cratering evidence from the surface of the Earth.

Asteroids assembled very early in our Solar System, approximately 4.6 Ga. Some asteroids such as carbonaceous chondrites provide a complex suite of organic compounds and water, which is biologically relevant to the origin of life. Although the early environment (including the constant bombardment by asteroids and volcanic activity) at the Eoarchean time was still hazardous to life, necessary ingredients for prebiotic synthesis were present in some form or another in impacting asteroids: liquid water, chemical building blocks, and some kind of

energy source. Water is the matrix and medium of life. It is a small, highly polar molecule so it is the ideal solvent for the chemistry of life—dissolving many molecules, transporting them to reaction sites, while preserving their integrity. It has a high capacity for absorbing energy, and has many properties that are essential for the functioning of proteins and cells. Ice and water vapor do not perform the same functions. Liquid water is considered essential to the initial development of life, because many chemicals brought by meteorites have parts that are attracted to it and parts that are repelled by it. Next to water, carbon is important for biogenesis because of its ability to form long chain-like molecules. Hydrogen, oxygen, nitrogen, phosphorous, and sulfur can bond with carbon in many ways, and large molecules made from these essential elements also tend to be stable. A small number of cosmic molecules is at the base of the cells, which are the starting point of life. These are lipids, nucleotides, and amino acids which are mandatory for the emergence of life.

Where life first arose on this planet is one of the keys to the discovery of how life originated. The young Earth probably had many sites in which the key biochemical requirements were met, such as the greenstone belts of Canada, Greenland, Australia, South Africa, and India. Hydrothermal systems have played important roles in shaping the volcanic–sedimentary sequences of the Archean greenstone belts, and habitats of early life [4]. Similarly, hyperthermophiles (superheat-loving bacteria and archaea) are considered the most primitive organisms from genomic sequences. Phylogenetic studies point towards the first cells occupying a hydrothermal vent environment [20]. Hyperthermophiles grow optimally above 80ºC and exhibit an upper temperature border growth up to 113ºC. Today, they are restricted to geothermally heated subterranean rocks, such as the boiling hot springs of Yellowstone National Park, hydrothermal impact crater-lakes, and submarine hydrothermal vents along the mid-ocean

ridge. Thus, we can reconstruct the early habitat of life from two critical lines of evidence: the likely habitats of the oldest organisms discovered from the widely scattered Archean greenstone facies of Canada, Greenland, Australia, South Africa, and India, and the recent ecology of hyperthermophiles. These two diverse lines of evidence – one from geological, the other from biological – suggest that the likely environment for life's beginnings was the steaming hydrothermal vents, full of volcanic chemicals and energy sources, making them as an ideal cradle for the emergence of life [21]. Hydrothermal conditions are conducive for the origin of life.

There are currently two competing hypotheses for the location of the primordial cradles of life: submarine hydrothermal vents [22,23], and terrestrial hydrothermal impact crater lakes [4,18,19,24]. Hydrothermal systems can develop anywhere in the crust where water coexists with a magmatic heat source. Hydrothermal sites are geochemically reactive habitats, where hyperthermophile thrives today around superheated water supporting chemosynthetic ecosystems. The chemicals found in these hydrothermal vents and energy they provide could have fueled many of the prebiotic reactions necessary for the evolution of life.

In recent times, debate rages over whether life began under the sea or on land. The submarine hydrothermal vents and terrestrial hydrothermal impact craters have been proposed as alternative sites for the origin of life. In both submarine and terrestrial hydrothermal settings, natural catalysts have facilitated reactions now performed by enzymes. In both sites, chemical reactions were far from an equilibrium state with a source of energy available in the vent environment that drove synthetic reactions. A prerequisite for prebiotic synthesis must be a process by which cosmic chemical ingredients could be sufficiently concentrated to undergo chemical and physical reactions. This question of concentration of biomolecules may settle

which early environment – submarine or terrestrial hydrothermal vents -- were likely incubators for life's beginnings.

The submarine hydrothermal vent hypotheses have considerable appeal but have not been universally accepted, partly because many aspects of the proposed scenarios remain unconstrained from geological evidence of the early Earth environment and chemical constituents of living cells. This theory suffers from the 'dilution problem' of the precise organic compounds. The cosmic ingredients would be dispersed and diluted, rather than necessarily concentrated, in the vastness of the global Eoarchean ocean, preventing them from assembling into the complex molecules of life [4,7].

One crucial precondition for the origin of life is that comparatively simple biomolecules must have the opportunity to develop into more complex molecules through the segregation and concentration of chemical compounds. In the open ocean, cosmic and terrestrial chemicals could never have mixed, concentrated, selected, or organized into more complex molecules. Moreover, many macromolecules found in the cell constituents, such as lipid cell membranes, RNA, proteins, and DNA, are all polymers and form via condensation reactions that need a fluctuating environment, which is sometimes wet and sometimes dry. Wet and dry cycling occurs every day on continental hydrothermal fields. This allows for concentration of reactants as well as polymerization. Submarine hydrothermal vent environments lack such wet and dry cycles for condensation [15].

Recent work suggests that the chemistry of modern cells still mirrors the original environment in which life first evolved. The chemical nature of terrestrial hydrothermal vents resembles the cell's composition of its cytoplasm more closely than the open ocean environment. Thus, protocells must have evolved in habitats with a high K^+/Na^+ ratio and relatively high

concentrations of Zn, Mn, and phosphorous compounds, as in terrestrial environments [25]. In contrast, seawater has 40 times more sodium than potassium, which may inhibit protocell formation. Geochemical reconstructions show that the ionic composition conducive to the origin of the cells could not have existed in marine settings but is rather compatible with inland geothermal systems, such as hydrothermal ponds and crater lakes. Between two terrestrial hydrothermal systems, crater basins retained heat for a longer period than ponds and could develop a network system, allowing many closely spaced craters with basins of different sizes to work in synchrony for biogenesis [4].

There has been growing evidence against the theory of submarine hydrothermal vent origins in recent years, which suggest that the lipid membranes and nucleotides could not have been formed in that environment during prebiotic synthesis. Amphipathic compounds, such as fatty acids, readily form membranous vesicles when dispersed in terrestrial aqueous phases. Recent experiments suggest that these vesicles, precursors of protocells, are stable in hydrothermal hot spring water but are unable to assemble in seawater. In particular, monocarboxylic acid vesicles are completely disrupted by low concentrations of divalent cations, such as magnesium and calcium, and by high sodium concentrations in seawater [26]. In the terrestrial setting, these vesicles can self-assemble and encapsulate nucleic acids such as RNA and DNA. These experiments clearly suggest that biogenesis started in terrestrial hydrothermal vent environments, not in the submarine hydrothermal vent setting [27]. Similarly, nucleobases delivered by meteorites and interstellar dust particles to early Earth could only be polymerized into RNA in terrestrial environments such as warm little ponds or crater lakes because of wet and dry cycles, a process that could not occur in the deep ocean [28].

There are also major tectonic problems with the submarine vent hypothesis. The hypothesis of the submarine hydrothermal vents as a likely cradle is difficult to explain in a one-plate Eoarchean Earth. How did submarine hydrothermal vents originate without plate tectonics? Today, they occur along or near the axis of the spreading ridge. But, since plate tectonics did not start before 3 to 2.5 billion years ago [29], there was no spreading ridge in the Eoarchean oceans for hydrothermal vent formation; we have to seek alternative hydrothermal systems on land.

Hydrothermal impact crater lakes on young crust appear to be the most compelling environment for the beginnings of life during the tail end of the Late Heavy Bombardment period [4]. Impact events on the Eoarchean crust produced thousands of craters on protocontinents resembling the surfaces of the Moon and Mercury [3]. But, unlike our planetary neighbors, the crater basins on the early continents were filled with water and cosmic building blocks, and there developed a complex system of subsurface hydrothermal systems that were crucial to the origin of life. The hydrothermal crater vents provide both chemicals and energy sources, making them feasible as a niche for the emergence of life. Impacts on a water-rich planet like Earth or even Mars can generate hydrothermal activity because of a high-energy, high-temperature event, creating underwater areas boiling with heat and spewing chemicals. Freshwater in the crater lakes was the ideal solvent for the cosmic ingredients to initiate prebiotic synthesis.

The hydrothermal crater lakes are geochemically reactive habitats, where hyperthermophiles (superheat-loving microbes) thrive today around superheated water. Hyperthermophiles are the most primitive living organisms, supporting the view that life began in extremely hot environments. On Earth more than 150 impact craters have been identified in a wide diversity of microbial habitats, ranging in size from the ~1.8 km diameter Lonar Lake

structure in India, to the ~250 km diameter Sudbury structure in Canada [30]. Of these, small crater-lakes with a central peak associated with hydrothermal systems (~5 km diameter) would be the modern analogs for the prime locations of prebiotic synthesis [31,32]. Most likely, closely-spaced craters were interconnected by networks of subsurface fractures, exchanging heat and chemicals for life-processing [4]. Hydrothermal impact crater lakes appear to be the most plausible environment for the beginning of life on early Earth, and, by analogy, on Mars. This is why NASA's Curiosity Rover is currently exploring the 3.8 billion-years-old Gale crater of Mars for evidence of early life [33].

A hydrothermal crater-lake provides different habitats, such as the highly fractured central uplift with spewing chemicals, impact ejecta deposits, annular basin, crater rim for sequestering the chemical reaction, post-impact water sediments in the lake floor, and temperature gradient of the water column, where a large number of chemical processes can occur simultaneously [18]. The cold freshwater of the crater lake is heated by molten rocks from the central peak that may reach high temperatures, driving the convection of lake water like a giant cooking pot. The cosmic building blocks, delivered by impacts, began to accumulate in the crater lakes, where the hydrothermal energy drove the synthesis of ever more complex organic compounds. The prolonged convective circulation of heated water, as well as the temperature gradient, allowed for the mixing and concentration of biomolecules and increased their chemical activities. Note, however, that not all the products of cosmic chemistry were used for the construction of life. Many molecules were discarded. Out of these vast number of diverse cosmic molecules, few would be selected for life-building processes, perhaps because of their molecular compatibility and cooperation to form more and more complex molecules. Complex molecules, such as RNA, protein and plasma membrane, must have originated from small molecules whose

reactivity was guided by physico-chemical conditions. Natural selection at the molecular level creating increasing complexity under prebiotically possible conditions emerges as a decisive factor in successive stages of biogenesis (Fig. 1). This may explain why life systems have a higher level of complexity than the non-living structures, and how few of those began to develop in the early stages of prebiotic synthesis.

Prebiotic chemical reactions in vent environments were the forerunner of the present-day metabolism. Throughout the living world, energy circulates almost entirely in the form of a single chemical entity, known as ATP (adenosine triphosphate). Hydrothermal vents produced a continuous stream of various chemicals and energy such as ATP, facilitating the chemical and catalytic reactions of cosmic ingredients [23]. ATP played an important role in primitive metabolism. Several key components of cells, such as lipid membranes, amino acids, sugar, phosphate molecules, and nucleotides, derived from meteorite impacts, began to interact in the crater lakes.

Mineral catalysts may have played important roles in establishing the early metabolism. Apatite might have helped in the building of the cell membrane because of its phosphate content. Metal ions of Fe, Mn, Zn, and Cu also were available in the vent environment, which help mediate catalysis. Crystalline surfaces of common rock-forming minerals such as pyrite and montmorillonite at the crater floor enhanced protometabolism by polymerization of nucleotides into RNAs and amino acids into peptides [1,4]. The metabolic activity of these early peptides became improved with the availability of phosphates.

The membrane is the defining boundary of a cell, the basic unit of life. Lipid cell membranes were first to form from cosmic ingredients by self-assembly, floating on the water surface of the lake like a thick oil slick (Fig. 2) [4,7]. These membrane-bounded components

began to encapsulate various biomolecules and monomers that separated them from their surroundings and enhanced the probability of interaction with one another. An early compartmentalization and development of protocells were necessary for the prebiotic synthesis. Compartmentalization offered many potential benefits to the emergence of protocells of increasing complexity, from simple amphiphilic vesicles to modern cell membranes.

<Figure 2 about here>

In the *chemical* stage, the combinatorial chemical reactions of the cosmic building blocks started to produce more complicated chemical species, which then autocatalyzed or catalyzed sets of more complicated reactions [34]. Various combinations of organic molecules were mixed and recombined to form complex interacting systems, then exposed to sources of energy such as heat, oxidation-reduction potentials, and ATP available in the hydrothermal vent environments [22]. This mixing and recombination probably did not occur in free solution but rather in fluctuating environments within primitive cell membranes [7,15]. All life today is homochiral and uses only left-handed amino acids (L-amino acids) and right-handed sugars (D-sugars). In the vent environment, both amino acids and sugars produced chiral molecules in 50:50 mixtures. In the early stage of chemical evolution, L-amino acids and D-sugars were selected by crystal faces of certain common minerals such as calcite, feldspar, quartz, and diopside from heterogenous populations [35]. The homochirality of biological molecules, L-amino acids and D-sugars is a signature of life. Chiral properties provide an extra dimension for molecular recognition, such as between the enzymes and their substrates, or between the amino acids and the RNAs.

The polymerization of chiral monomers, such as amino acids and nucleotides is the next key event in the chemical evolution. The polymers, such as peptides and RNAs, are central to

life processes. Condensation or loss of water (dehydration) between two monomer molecules facilitates linking them into long chains of polymers. A repeated wet-and-dry cycle on the surface of the crater basin facilitated the polymerization of these monomers [15]. The catalytic nature of the mineral surface of the crater floor also might have played an important role in polymerization of RNA and peptide molecules. Several mineral catalysts for polymerization, such as montmorillonite clay and pyrites, abounded at the hydrothermal vent environments [35,36]. Empty lipid membranes began to encapsulate various monomers and polymers for efficiency, stability, and molecular symbiosis [6,15].

The RNA world model has become the main paradigm in the current origin of life research in which RNA assumed informational and functional roles [37-41]. In our previous paper [41], we have argued that peptide/RNA world is more parsimonious than the RNA world for biogenesis. Peptides were relatively easy to synthesize in the prebiotic environment than RNAs, and there is no justification for excluding peptides from prebiotic chemical reactions. Both ribozymes and peptide enzymes, when worked together, could catalyze chemical reactions more efficiently than ribozymes alone. The peptide/RNA world is supported by several authors in recent times [42-48]. These authors have argued that a single polymer like RNA could not carry out all of the necessary processes required for the emergence of the genetic code and protein synthesis. Various forms of peptides, which were available in the prebiotic vent systems, were also essential to build the translation machineries and genetic code to carryout necessary processes for protein synthesis. A lonely RNA world could not achieve a similar feat without the active collaboration of peptides.

A reciprocal partnership between peptides and RNAs were essential initially for the beginning of the *information stage*; both contributed rudimentary information coding and

catalytic rate of accelerations. The peptide/RNA world scenario documents a path to increasing complexity that created translation and the genetic code. We suggested that the demand for a wide range of protein enzymes over peptides in the prebiotic reactions was the main selective pressure for the origin of information-directed protein synthesis [42]. Once the programmed protein synthesis was available in the prebiotic system, the peptide/RNA world gave rise to the protein/RNA world. A wide range of protein enzymes and structural proteins were synthesized at this stage by translating genetic code; some of these enzymes catalyzed more and more complex biochemical reactions.

The second stage of metabolism from pre-RNA metabolism, defined as the first set of reactions catalyzed by protein enzymes (and perhaps, ribozymes), prefigured present-day metabolism, and perhaps already included certain central systems such as the glycolytic chain and the Krebs cycle [42,63]. Centrally located within this network are the sugar phosphate reactions of glycolysis and the pentose pathway. This stage of metabolism appeared in the peptide/RNA world and modified and refined continuously during the origin of the proteins leading to the first cells. The optimization of the genetic code and the development of the protein synthesis were central to the emergence of metabolism. As more enzymes were added and started to build their own network, new pathways could have developed.

In this paper, we discuss the evolution and functional repertoire of translation proteins that led to hierarchical emergences of viruses, plasma membranes, cytoplasm, DNA, and finally the first cells.

2. Encapsulating the Molecules of Life in the Peptide/RNA World

Cell membranes were essential to the development of life forms in the prebiotic Earth. The primitive cell membranes enclosed the life-building molecules, defined the boundaries from environments, and enhanced chemical reactions inside protocells. Each cell membrane was a very thin film of lipid molecules, about 5 nm thick and was a dynamic fluid structure. The origin of cellular life presumably occurred by self-assembly of organic compounds into encapsulated systems capable of catalyzed polymer synthesis [50,51]. Thus, the availability of a primitive cell membrane component in a hydrothermal vent environment was a prerequisite for biogenesis. The cell membranes are built of components that have a remarkable capability of spontaneous self-assembly. The chemical requirements for formation of cell membranes from individual molecules are remarkably simple. The lipids that form the lipid bilayer common to all biological membranes are not gene products. Therefore, the development of genetic material need not have to precede the development of the membrane. However, the development of membranes could have facilitated the development of self-replicating genes by providing them a protected space in which to evolve and eventually function.

The lipid bilayer, a two-dimensional fluid, has been firmly established as the universal basin of all cell membranes. All modern cell membranes are composed primarily of diacyl or dialkyl glycerol phospholipids; the first cell membranes are thought to have assembled from simple, single-chain lipids from cosmic ingredients [52]. Primitive cell membranes were the first macromolecules among cell components that appeared in the *geological* stage in the hydrothermal crater vent environment from the cosmic ingredients (Fig. 2). Hydrocarbons and their primitive derivatives were available in the prebiotic environment to synthesize primitive membranes.

In the peptide/RNA world, the primitive cell membrane or protocell provided a way to keep peptide and RNA molecules as well as translation machinery together at relatively high concentration [42]. The primitive membrane created an internal environment within which RNAs can reside and metabolic activities can take place without being lost to the hydrothermal vent environment. This compartmentation opened the opportunity for a different chemistry inside compartment than that of the outside compartment. Such compartmentation is a necessary prerequisite for maintaining the integrity of cooperative molecular systems that are necessary for metabolism and protein synthesis. It separated these reactive molecules from its environment for further concentration and chemical reactions. A great milestone in the origin of life occurred when cooperation of macromolecules occurred within a primitive cell membrane. The concentration of molecular building blocks was facilitated by enclosure within a small volume of the primitive cell membrane. Such a membrane allowed encapsulated polymers to remain together and allowed information transfer between molecules within a protocell. Encapsulated systems of molecules would be essential for life to begin [15]. Membrane-enclosed molecular cooperatives preceded the first cells.

With the availability of RNA-directed proteins, there was a gradual evolution of the constituents of cell membranes in the protein/RNA world. Lipids and proteins are the major components of a modern cell membrane, although proportions of lipid and protein vary widely. Lipids constitute the bulk of the membrane; they establish the physical integrity of the membrane and create an effective barrier to the rapid passage of hydrophilic materials such as water and ions. Three classes of lipids—steroids, phospholipids, and fats are the three most important types found in modern cell membranes. Isoprene and fatty acids are building blocks for these three types of lipids (Fig. 4).

The permeability of lipid bilayers can be altered radically by transmembrane spanning proteins, which are scattered throughout the plasma membrane. They make up about 50% of the plasma membrane. These proteins are responsible for the passage of ions, polar molecules, and large molecules that do not readily cross the phospholipid bilayers on their own. Membrane-mediated proteins provide a selective permeability that acts as channels, carrier, and active transporters (pumps). The phospholipids and proteins move back and forth within the plasma membrane, making the plasma membrane a fluid structure. Such membranes are extremely permeability barriers, so that modern cells have complete control over the uptake of nutrients and export of wastes through the specialized channel, pump, and pore proteins embedded in their membrane.

<Figure 4 about here>

It seems unlikely that phospholipids, sterols, and proteins contributed to the first forms of protocells because they are products of complex and highly metabolic pathways that incorporate multiple enzyme-catalyzed steps. Most likely, the earliest membranes were composed of simple amphipathic lipids, composed of hydrocarbon derivatives between 10 and 20 carbon length with carboxylate or hydroxyl head groups. These primitive membranous structures were constructed from single chain fatty acids, fatty alcohols, and monoglycerids. The most convincing aspect of this argument is that prebiotic availability of such amphiphiles has been established from carbonaceous chondrites and interstellar dust [52].

Modern cell membranes are composed of complex mixture of amphipathic mixtures molecules such as phospholipids, sterols, and many other lipids as well diverse proteins that perform transport and enzymatic functions. Such membranes are extremely good permeability barriers so that modern cells have complete control over the uptake of nutrients and export of

wastes through the specialized channel. Modern cells therefore require sophisticated protein channels and pumps to mediate exchange of molecules [53]. We suggest that membranes evolved in two stages from ancient lipid membranes in the protein/RNA world: the development of phospholipid membranes, followed by the evolution of plasma membranes.

2.1. Primitive Amphipathic Cell Membranes

The bilayer membranes that surround all present-day cells and act as boundaries are thought to have originated in the spontaneous self-assembly of amphipathic molecules into membrane vesicles [15]. The first cell membranes were assembled from simple, single-chain lipids, such as fatty acids and their derivatives, which were available in the prebiotic environment. These amphipathic molecules have polar and nonpolar groups among the same molecule and spontaneously self-assemble into lipid layers when mixed with water. In the absence of self-assembly processes, modern cell membranes could not have been evolved. The hydrophilic head is oriented into water and the hydrophilic tails are attracted towards each other to expel water. The lipids can form vesicles out of a monolayer of molecules or out of a bilayer of molecules. In a monolayer micelle, the external surface of the vesicle is always the hydrophilic end, whereas the inside of the vesicle contains the hydrophobic portion. This type of vesicle can only trap oils, not water, and therefore could not lead to the production of a cell. On the other hand, a vesicle formed from a bilayer has a hydrophilic surface on both the exterior and interior surfaces. Such a vesicle can trap water and water-soluble molecules like peptides, ribozymes, RNAs, sugars, and proteins (Fig. 4). The bilayer membranes are stabilized by the hydrophobic effect and van der Walls interactions [52]. This property of self-assembly is by no

means rare. Even fatty acids on their own, such as long-chained carboxylic acid extracted from meteorites, forms membranous vesicles in water. Similarly, phospholipid membranes self-assembled to vesicles in water. Both fatty acids and phospholipids form vesicles of similar size; both are thermally stable; and both are similar in tensile strength; and have similar permeability selectivity trends for small, uncharged solutes [54].

<Figure 4 about here>

The secret of membrane construction is the lipid bilayer. The simplest example of lipid bilayer is the soap bubble. However, the lipid membranes are microscopic and more complex than soap bubbles. Individually, lipids are tiny molecules, but when grouped together, they form the largest structure of the cell. When placed in freshwater, lipid molecules aggregate to form huge waterproof sheets. These amphiphiles assemble spontaneously into closed membranous vesicles. The unusual interaction of lipids with water makes them so useful. They are composed of a small hydrophilic 'head' connected to one or two long hydrophobic 'tails.' When placed in water, lipid molecules spontaneously aggregate, packing molecules side-by-side to shelter the long tails away from the water.

Packing constraints imply a minimum chain length of fatty acids required to form vesicles, about 10 to 20 carbons long [52,55]. Primitive cell membranes needed a membrane compartment for many of the same reasons that modern cells do: to keep molecules that are important for cellular growth and survival readily accessible, and to keep unneeded or potentially harmful molecules outside the protocell.

Carbonaceous chondrites contain a rich mixture of organic compounds including a variety of membrane-forming lipids that were synthesized abiotically in the early Solar System. These carbonaceous chondrites may have brought these fatty lipid molecules along with other

organic compounds to the early Earth during early impacting processes [56]. Organic compounds extracted from the Murchison meteorite form vesicles in aqueous solution mimicking cell membranes. These spheres are not true cells. They are empty, but they are composed of heterogenous mixtures of amphiphiles consisting of fatty acids, fatty alcohols, and monoglycerids forming bilayer lipid molecules on their surfaces. They have enough mechanical strength to have contained other cell components, such as nucleic acids and proteins, and to have given them some protection from the external environment [56]. Laboratory simulation of interstellar ice mixtures produce amphipathic vesicle-forming compounds similar to those found in the Murchison meteorite [57]. Similarly, synthesis of primitive membranes in a simulated hydrothermal vent environment produced heterogenous mixtures of amphiphiles and fatty acids as products of Fischer-Tropsch-type reactions [58]. Thus, the study of carbonaceous material and laboratory models show that vesicle-forming amphiphiles likely were present on the early Earth, and played the critical role for boundary membranes required for early protocells. Fatty acid vesicles may be further stabilized by the admixture of other simple amphiphile such as fatty alcohols and fatty acid glycerol esters. Moreover, short-chain amphiphile-based vesicles have properties similar to those of liposomes formed from phospholipids that are primary components of modern cells [59].

Single-chain lipids, such as fatty acids, are thus attractive constituents of a protocellular membrane because they are chemically simple, common cosmic ingredients, and prebiotically possible. Moreover, they readily form bilayer membranes in water. Fatty acids, readily form membranous vesicles when dispersed in terrestrial aqueous phases. These vesicles may be further stabilized by the admixture of other amphipathic molecules such as fatty alcohols and fatty acid glycerol esters. Recent experiments suggest that these vesicles, precursors of protocells,

are stable in hydrothermal hot spring water but are unable to assemble in seawater. In the terrestrial hydrothermal setting, these vesicles can encapsulate nucleic acids such as RNA and DNA. These experiments clearly suggest that biogenesis started in terrestrial hydrothermal vent environments, not in the submarine hydrothermal vent setting [27]. Here, clay particles such as montmorillonite at the mineral substrate at the crater floor converted the fatty acid micelles into vesicles. The other favorable quality of terrestrial hydrothermal sites is that they are fluctuating environments undergoing continuous cycles of hydration and dehydration. The hydration/dehydration cycle permits encapsulation and also promotes polymerization [15,60]. The formation of the membrane was crucial to the development of cellular life. The lipid cell membrane was probably the first cell component to emerge from cosmic ingredients in the geological stage, and acted as the interface between organic molecules and the environment (Fig. 2). These lipid cell membranes offer some flexibility and self-sealing and encapsulated various monomers, polymers, and other organic compounds available in the vent environments. These protocells began to develop several functional properties, including self-assembly of boundary membranes, transport of monomers, and encapsulation of polymer systems capable of growth and of developing an information system.

Fatty acids are fundamental to the nature of primitive cell membranes not only for their chemical simplicity, but for their dynamic properties that are essential for membrane growth and permeability. Once formed, fatty acid vesicles are highly stable, and appear upwardly unchanging over the course of days or months. At the molecular level, however, fatty acids are extremely dynamic, entering and exiting the vesicle bilayer, as well as flipping between the inner and outer leaflet of the membrane. Fatty acid flip-flop may be responsible for increased permeability. Membranes consisting of fatty acids are reasonably permeable to small polar

nutrients such as nucleotides and even to charged species such as ions [53,54]. Moreover, dynamic fatty acid membranes are permeable to nucleoside mono- and diphosphates and are necessary for spontaneous growth and division [59]. Prebiotic vesicles were certainly composed of complex mixtures of amphiphiles. Membranes composed of mixtures of amphiphiles often have superior properties such as stability and tolerance over a wide range of pH and ionic conditions to those composed of only single species such as fatty acids [26]. The fabric of all modern membranes is the lipid bilayer, a bimolecular layer typically made of isoprene derivatives. The long-chain acids and alcohols that contribute the amphipathic property of contemporary membrane lipids are another possible component of prebiotic membrane structure [10]. These early cell membrane systems exhibited many of the characteristics that modern biological membranes possess without relying on genetically encoded transport systems.

Fatty acids are attractive as the likely prebiotic membranes in that they are chemically simpler than phospholipids and were readily available in the vent environment. The lipid bilayer is an inert structure that can neither mediate nor control the multiple exchanges of matter and information that must necessarily take place between the protocells and their environment. Once the catalytic species such as peptides and ribozymes were incapsulated in fatty acids vesicles, access to energy sources and nutrients became essential. Since early primitive vesicles lacked a specialized transport system, then simple uptake mechanisms, such as passive diffusion, would play an essential role in the nutrient and energy uptake across boundary membranes. In hydrothermal vent environments, both nutrients and chemical energy including ATP were available [23]. Lipid bilayer membranes offered an important role in energy transduction. Transmembrane diffusion of the lipid cell membrane was fast enough to keep up with the

demands of a primitive metabolism [59]. Fatty acid vesicles are relatively permeable to ionic and polar solutes.

Vesicles formed by fatty acids have long been studied as models of protocell membranes. Fatty acids membranes can be made to grow and divide under laboratory conditions, and thus provide a model system relevant to the emergence of cellular life [62]. The primitive cell membranes developed some kind self-reproduction process. Self-reproduction requires not only replication of genetic material and catalysts, but also the production of additional membrane surface to accommodate growth and capability of producing daughter cells. RNAs must code for three essential components of the protocell: the ribozymal polymerase activity is required for transcription of the genes into other active ribozymes, including replication of the genetic RNA itself; a ribosomal acyl transferase activity synthesizes new membrane molecules from their precursors; and finally, an RNA fragment is required to trigger the creating of compartment boundaries [59].

Protocells used both self-assembly and directed assembly processes to grow and evolve. Self-assembly was essential to the synthesis and stability of membrane structures and protein folding. Directed assembly underlies the synthesis of proteins according to the base sequences of mRNA. In the beginning of the information stage, the translation system and the genetic code evolved step by step in the peptide/RNA world in the enclosed cell membranes [42]. Once the mRNA-directed protein synthesis became established, various enzymes were created to meet the demand of catalysis and metabolism. The next step involved encapsulation of a more complex enzyme system capable of catalyzing fatty acids to form phospholipids. The transition from single-chain lipids to phospholipids was triggered by the availability of a wide range of enzymes

in the protein/RNA world that catalyzed the conversion of lipid membranes into phospholipid membranes.

3. Protein/RNA World

In our previous paper [42], we have discussed in detail the origin of the prebiotic information system in the peptide/RNA world that permitted the subsequent evolution of the genetic code, translation machine, and protein synthesis through a series of transitional stages. Proteins were assembled from amino acids using information encoded in mRNA and translated to protein by tRNA. The ensuing peptide elongation was catalyzed by rRNA in the ribosome. Each protein had its own unique amino acid sequence of the mRNA gene.

Proteins control most of the functions of a cell, breaking down nutrients, assembling cellular components, copying DNA, and so on. They truly occupy a central position in the organization of the first cell. They act as enzymes that permit only a few of the many possible reactions among cellular components to take place. They form channels in plasma membranes, allowing specific substances to enter and leave, while excluding others. Protein molecules owe their properties to their three-dimensional shapes, which are themselves determined by their amino acid sequences of their constituent chains. These properties in turn determine how a protein biologically functions: whether it will bind certain organic molecules and catalyze their reactions or form regular structure such as a helix and act as a building material.

The first proteins were most likely short, about 25 amino acids long. Protein molecules of this short length displayed enzyme-like activities. In contrast, many modern-day proteins contain several hundred amino acids. Most likely, these long molecules arose by gradual lengthening of

first products of protein synthesis [63]. Proteins are the primary functional biomolecules of life. Once formed, proteins performed a vast array of functions during biogenesis including catalyzing metabolic reactions and reinforcing cell membranes. The overwhelming number of efficient proteins in the biochemical synthesis occurred in the protein/RNA world. With their astonishing versatility, protein enzymes would have taken ribozyme's role in assisting genetic copying and metabolism. Different kinds of enzymes were in great demand in the protein/RNA world. In order to be useful, an enzyme must necessarily have specific substrate available in its environment on which to act. The enzyme must also have an outlet for the products it forms. These substrates and outlets must have been provided by the primitive metabolism in the vent environment that supported the protocells at the time [63]. These newly formed enzymes carried out hundreds of chemical reactions that took place in the protocell. Structural proteins, on the other hand provided structure and support for the cell membrane.

Four major events followed after the availability of template-directed proteins but with considerable overlap. These affected, first, the efficiency of the translation machinery, then, resilience of the coding system, and finally, the quality of the synthesized proteins. These protein-mediated events are: (1) the transition from the phospholipid membrane to the plasma membrane; (2) the origin of prebiotic cytoplasm; (3) the beginnings of the virus world; and (4) the advent of DNA. The newly synthesized protein enzymes helped to catalyze and mediate these critical molecular evolutions, favored by strong selective forces.

4. The Evolution of Protocellular Membranes

The phospholipid bilayers were impermeable to most water-soluble molecules. This property made bilayers excellent boundaries that allowed for protocells to maintain an internal composition different from that of the surrounding medium. But protocells could not survive and evolve sealed off from the outside. They must be able to take up nutrients, got rid of waste products, and respond to environmental signals. These functions were carried out by proteins inserted into bilayers, improving the bilayer's permeability.

4.1 The Origin of the Phospholipid Membrane

New phospholipid (more accurately called glycerophospholipid) membrane was made initially in prebiotic environment by a gradual modification of the existing fatty acid membrane. The synthesis of phospholipid requires an activated intermediate, phosphatidate acid (diaglycerol 3-phosphate). As discussed earlier, three critical molecules—fatty acid, glycerol, and phosphate -- were available in the hydrothermal vent environment for synthesis of the phosphatidate acid by non-enzymatic synthesis [65]. The phospholipid is synthesized by a series of enzyme-catalyzed energy dependent reactions from the phosphatidate acid (Fig. 5). This was possible when RNA-directed protein enzymes were available to convert the phosphatidate acid to phospholipid. As such, phospholipid provides a marked contrast with fatty acid as a membrane component. Because of its molecular complexity, it is generally believed that the prebiotic synthesis of the phospholipid was difficult [59]. More recently, several studies have achieved the synthesis of the phospholipids and related compounds, such as acylglycerol and glycerol phosphates, suggesting that such molecules may have been present in the prebiotic environment in trace quantities [11, 26, 64-66]. Moreover, abiotic formation of the ester bond among lipid compounds, including

acylglycerides, is possible under simulated hydrothermal conditions, provided the precursors that are present are at sufficient concentrations [67].

Most likely, the phospholipids did not become available in the prebiotic environment until various protein enzymes were available and metabolic pathways for their catalyzed synthesis evolved. Non-enzymatic synthesis of ester bonds to produce monoglycerides might have been the first step toward the glycerolipids [65]. The phospholipids require only three precursors: fatty acid, glycerol, and phosphate (Fig. 5A). Abiotic synthesis of aliphatic lipids, fatty acids, and acylglycerols has been reported to occur at elevated temperatures and pressures under simulated hydrothermal conditions [58]. With the biosynthesis of proteins, many critical enzymes were available for the synthesis of the phospholipids.

The synthesis of the phospholipids requires an activated intermediate. The prebiotic pathway from fatty acids to the simplest phospholipid, phosphatidic acid, occurs via successive acyl- and phosphotransfer reactions. The first step in the synthesis of the phospholipids is the synthesis of phosphatidate acid (diglyceride 3-hosphate), which is formed by the addition of two fatty acids to glycerol 3-phosphate. Here we explore the non-enzymatic pathways of the emergence of the phosphatidate in the peptide/RNA world. From three building blocks—fatty acids, glycerol, and a phosphate ion, which were available in the prebiotic environment [65], we show the gradual evolution of monoglyceride, diglyceride, and triglyceride by the condensation reaction. When glycerol and fatty acid react, a water molecule is expelled, forming an ester linkage. The production of diglyceride is considered first, since it will lead directly to the biosynthesis of the phosphatidate. Diglyceride is formed when the glycerol and fatty acid chains become joined by two ester linkages. Non-enzymatic synthesis of ester bonds to produce the diglycerides might have been the first step toward glycerolipids [65-67]. The phosphorylation of

diglyceride, in turn, would give rise to the phosphatidate acid, which is essentially a diglyceride in which a phosphate group has been added to a single glycerol molecule (Fig. 5C, 5D).

The non-enzymatic synthesis of the activated phosphatidate acid was at a pivotal point in the lipid biosynthetic pathways. It served as the precursor for the formation of the glycerophospholipid (commonly called phospholipid) membrane by enzymatic synthesis. Glycerophospholipids are the main constituents of membrane bilayers. Enzymatic synthesis pathways evolved over time when RNA-directed protein enzymes were available in the protein/RNA world. The phosphate group of the activated phosphatidate acid is esterified to an alcohol to produce a variety of phospholipids including the attachment of choline, ethanolamine, serine, and inositol to the phosphate group of phosphatidic acid. Names of phospholipids then include phosphatidylcholine (phosphate + cholin), phosphatidylethanolamine (phosphate + ethanolamine), phosphatidylserine (phosphate + serine), and phosphatidylinositol (phosphate + inositol). If the alcohol is choline, the product is phosphatidylcholine. Of these, phosphatidylcholine is the most abundant phospholipid in cell membranes. Different modifiers give the phospholipids different properties and roles in a cell. Three successive enzymatic methylation could convert the phosphatidate to phospholipid.

<Figure 5 about here>

A phospholipid consists of a polar headgroup on one end of the molecule and fatty acids chains on the end. These chemical structures create an amphipathic liquid. In solution, they instantly form bilayers that are selectively permeable. The phospholipids are composed of a polar head group (usually a negatively charged phosphate group and glycerol); it is hydrophilic. Phosphate is a primary anionic component of most phospholipid membrane lipids (Fig. 5). The phospholipid tails consist of two long fatty acid chains, which are hydrophobic and avoid

interactions with water. Two fatty acids are attached to a glycerol by ester or other bonds. The polar head group and fatty acid chains are attached by a 3-carbon glycerol unit [15].

The phospholipid molecules have a hydrophilic head end and two hydrophobic tails that will not mix with water and will avoid being surrounded by it (Fig. 6A). Because these amphipathic molecules have both a hydrophilic and hydrophobic group on the same molecule, they can undergo self-assembly into a cell. In an oil slick, the hydrophobic tails mix with oil while the heads stay close to the water in a monolayer cell (or micelle). When placed in water, the phospholipids will orient themselves in a bilayer in which non-polar tail regions face the inner layer of the bilayer (Fig. 6B). Being cylindrical, the phospholipid molecules contribute structural stability and create a semipermeable environment. The same forces that drive the phospholipids to form bilayers also provide a self-healing property. Admixture of cholesterol helps to stabilize the bilayer.

<Figure 6 about here>

Modern phospholipid-based cell membranes are formidable barriers to the uptake of polar and charged molecules ranging from metal ions to complex nutrients and nucleotides. They require special protein transporters to allow their passage through the membrane. The phospholipid membranes are stable under a wide range of temperature, pH, and salt concentrations. A recent experiment suggests that the phospholipid membranes can self-assemble and are surprisingly permeable to transport molecules across its membranes [53].

The evolution of phospholipid membranes was a critical and necessary step for the early evolution of cells. The transition from single-chain lipids to double-chain phospholipids had to be gradual. Recent experiments suggest that low levels of phospholipids could drive the growth of fatty acid vesicles by competition of monomers with neighboring vesicles lacking the

phospholipids. This competitive growth would have provided strong selective advantage for primitive cells to evolve the catalytic machinery needed to synthesize the phospholipids from their single-chain precursors. Growth results from decreasing fatty efflux from the membrane with increasing phospholipid content, suggesting an evolutionary arms race among primitive protocells [64]. The gradual transition from a fatty acid bilayer membrane to a phospholipid bilayer membrane is shown here in Fig. 7.

<Figure 7 about here>

Increasing phospholipid content inhibits the permeability of fatty acid membranes through changes in bilayer fluidity. The emergence of the phospholipid membrane would have therefore imposed new selective pressures for the evolution of a more permeable plasma membrane with the protein transport system. Proteins that were amphipathic could be inserted into the phospholipid bilayer to increase permeability.

4.2 The Origin of the Plasma Membrane

Here, we suggest the likely scenario for the origin of the plasma membrane from the phospholipid membrane (Fig. 8). A new class of proteins emerged in the protein/RNA world that played critical roles for the conversion of the phospholipid membrane to the plasma membrane. Proteins could be amphipathic because they are made of amino acids, and amino acids have R-groups that range from highly nonpolar to highly polar. To incorporate in the phospholipid membrane, the nonpolar amino acids would be selected in the interior of the lipid bilayer, while the polar would be selected alongside the polar heads of the surrounding water.

As the phospholipid membrane began to interact with new generated amphipathic proteins by translation, endosymbiosis would integrate these proteins to the phospholipid membrane (Fig. 8A). These protein symbionts would stabilize the walls of the membrane, allowing them to resist disruptive forces such as the mechanical shear caused by the convection current of the vent. The resulting increase in osmotic pressure and membrane tension would create a driving force for an increase in the membrane area, therefore encouraging symbiosis between proteins and lipid membranes that triggered the origin of the primitive plasma membrane for stability to prevent a burst of protocells (Fig. 8B). This is how more sophisticated plasma membranes might have formed by endosymbiosis to make them more permeable. The plasma membrane of modern cells is composed of roughly equal parts proteins and lipids by weight. Most likely the primitive protocells had a higher percentage of lipids than proteins from the initial endosymbiosis that favor high degree of thermostability. The plasma membrane acted as a selectively permeable barrier, preventing some substances from crossing while permitting other substances to enter and leave the protocell. The selective permeability of the plasma membrane and the specificity of transport proteins made it possible to create an environment inside the protocell that was radically different from the prebiotic soup and amenable to biogenesis.

The first step of survival in confinement was the possibility for the protocells to take in nutrients and energy from the outside environment and get rid of waste material. The simplest way in which fully enveloped protocells could fulfill this condition was by means of pores, mere holes kept open in the phospholipid bilayers by some kind of inserted protein framework. The insertion of proteins in bilayer lipid vesicles was an essential first step that affected the phospholipid bilayer's permeability, facilitating transport and other molecules into the protocells.

In primitive membrane transport, passive transport was used when molecules were moved across the plasma membrane of the protocells along, down, or with their concentration gradients. The molecule would flow from where it was at higher concentration to where it was at a lower concentration.

Next came transport facilitators, which were transmembrane proteins that act as restricted passages for certain specific substances. A more sophisticated kind of molecular passage was the gated channel-like facilitators that let certain substances of a given chemical specificity to move through passively but they were unidirectional and regulated by a gate that needed to be unlocked by some channels or electrical signal.

The next improvement in the building of molecular transport systems was active transport, where molecules were moved against their concentration gradient in energy-requiring processes and the machinery involved was correspondingly more complex. The systems that carry out such active transport are called pumps. The energy used was derived from ATP, the universal currency of energy. There were various ATP-powered pumps that were used to transport ions and molecules against their concentration gradients. The plasma membrane would have played an essential role in the generation of metabolic energy, and transformed it into useful ATP.
The classical fluid mosaic model of the structure of the cell membrane distinguishes between two types of membrane proteins: peripheral proteins and integral proteins; the former occurs only outside the lipid bilayer, while the latter spans the entire membrane for transport of ions and molecules [68]. The phospholipids and the plasma membranes make back and forth movements within the plasma membrane, making the plasma membrane a fluid structure. The key point is that the arrangement of proteins makes the interior and exterior surfaces of the plasma membrane very different (Fig. 8C). Thus, the plasma membranes are a mosaic of the phospholipids and the

different types of proteins, and the overall structure is dynamic and fluid. In protocells, the membrane proteins were responsible for the passage of ions, polar molecules, and large molecules that did not readily cross the phospholipid bilayers on their own.

The permeability of the phospholipid bilayer was altered radically by membrane transport proteins, which were scattered throughout the plasma membrane. Selection pressure favored the evolution of the plasma membrane over the phospholipid membrane to overcome the reduced membrane permeability. The integral proteins would allow only certain molecules to enter the protocell. In this way, the protocell could fine-tune the selection of what got in and what did not. The peripheral proteins, on the other hand, acted like sensory antennae, making it possible for protocells to gain information about their immediate environment. In addition to transporting substances into and out of the protocell, the plasma membrane began to function in energy production. It created an internal environment that was more conducive to life synthesis than the external environment. Several layers of the plasma membrane were added with the gradual evolution of the protocells to the first cells.

<Figure 8 About here>

Proteins are responsible for most of the dynamic processes carried out by the cell membrane, including the transport of molecules into and out of the cell (Fig. 6C). The plasma membrane separates the cell from its environment and is selectively permeable: it chooses what enters and exists in the cell. Receptor proteins are the gatekeepers; they detect signals from the environment of the cells; the transport proteins help some molecules get across the membrane. Certain membrane proteins act as enzymes. The plasma membrane was the ideal microenvironment to experiment with synthesis of more complex nucleic acid such as DNA for the permanent storage of the information system.

5. The Origin of Cytoplasm

Once different kinds of proteins were synthesized by translation machinery and available for prebiotic synthesis, they began to accumulate in the primitive cytoplasm inside protocells. This primordial cytoplasm became the ready source of a variety of proteins, when needed. The simple aqueous solution (prebiotic soup) inside the protocell was gradually converted to a viscous, gel-like cytoplasm that increased the protocell volume and provided some rigidity of its spherical shape. The first cytoplasm in protocells and the semipermeable plasma membrane from the phospholipid membrane most likely organized at the same time with the availability of proteins (Fig. 8D). The prebiotic cytoplasm provided a stable microenvironment for the organization of all nucleic acids, lipids, enzymes, proteins, other macromolecules, molecules, ions and ribosomes, as well as water and salts that were all encased in the plasma membrane. This primitive cytoplasm became a complex, crowded system containing a wide range of molecules – from ions and small molecules, to macromolecules like proteins, nucleic acids, and ribosomes. Many metabolic reactions, including protein synthesis, and the transition from RNA to DNA began to take place in this primordial cytoplasm. Over half of the molecules were actively involved in the synthesis of proteins. Some of these proteins were used in the synthesis of viruses and DNA molecules, while others were engaged in energy production. The constituents of the cytoplasm were moved across the protocell depending on their requirements. Primitive cytoplasm supported and suspended these molecules in its gel-like substance. This primordial cytoplasm became the site for most of the enzymatic reactions and metabolic activity of the protocell. The primitive cytoplasm was confined to the outside by the plasma membrane,

the latter began to regulate the passage of some substances, such as organic molecules, ions and water, preventing the passage of some other substances to maintain the content of the primitive cytoplasm. Other compounds moved passively across the membrane.

Two major innovations took place in succession, exploiting these new protein reserves in the prebiotic cytoplasm: (1) the origin of the virus world; and (2) the advent of the DNA world. Here we discuss the origin of the primordial viruses in the vent environment that gave rise to the DNA world.

6. The Beginnings of the Virus World

Viruses straddle the line between living and nonliving. They are tiny, noncellular, microscopic parasites that infect virtually every type of known cell. Even simpler and smaller than a bacterium, a virus has a diameter of 20–400 nm. Because they are not living organisms in a true sense, they require the biochemical machinery of a cell to reproduce. A virus is nothing more than a few strands of genetic material wrapped in a package of protein—a parasite, unable to function on its own. In order to survive, it must find a cell to infect. Only then can the virus take control of the host's cellular machinery and use it to churn out thousands of copies of itself. These viruses then move from one cell to the next, transforming each new host into a factory that makes even more viruses.

Their hallmark characteristics, namely their small size, tiny genomes, and parasitic dependence on cellular hosts for reproduction, set them apart from all other living things despite their animation. However, the discovery of the giant viruses (> 400 nm), called the Mimivirus, with massive genomes and the most complete resources for building proteins further blurs the

established boundaries between viruses and the smallest parasitic cellular organisms. The simple size-based distinction between viruses and cells is no longer tenable. However, its icosahedral ultrastructure of capsid coat, and its typical eclipse phase in its life cycle, support the viral nature of the Mimivirus. Furthermore, the Mimivirus lacks universal bacterial genes, such as encoding ribosomal RNA and proteins [69].

Viruses can be defined as capsid-encoding organisms as opposed to ribosome-encoding cellular organisms [70,71]. Viral particles are by far the most abundant biological entities on our planet, greatly outnumbering all their cellular hosts put together; most of the biomass in the ocean is made up of viruses [72]. The genetic diversity of viruses is enormous as well, in part because they can acquire genomes from their hosts and they can later paste these genes into new hosts. The viruses are agents for gene dissemination, evolution, and biodiversity.

The simplest viruses have just two components: a nucleic acid core and an outer protein capsid shell. The genome, which may be DNA or RNA, contains the instructions for taking over cells, making capsids, and creating more virions, or viral particles. There are many types of viruses, classified by their size and shape, by their genetic material (RNA or DNA), and by their host organisms. The majority of viruses have a genome based on DNA, although a significant minority has RNA genes. Viruses come in three common shapes: helical, polyhedral (such as icosahedral), and complex viruses, the latter often possess unique structure or extension on virions. Viruses are highly diverse in their morphology and in the nature of their genetic material. The genomes may be single-stranded RNA (ssRNA), double-stranded RNA (dsRNA), single-stranded DNA (ssDNA), or double-stranded DNA (dsDNA). Encapsidation of viral genomes constitutes a virus particle.

In their overall structure, viruses fall into two categories: enveloped and non-enveloped. Non-enveloped viruses have an extremely simple structure. They consist of genetic material and possibly one or more enzymes that are encased inside the capsid. Enveloped viruses are more derived and complex, where the capsid is surrounded by an envelope. The envelope consists of a phospholipid bilayer with a mixture of viral proteins and proteins derived from the plasma membrane of the host cell [73,74].

Viruses are completely selfish [75]. They can't reproduce on their own, because they lack ribosomes and the rest of the living cell's protein-making machinery. They can only reproduce by invading host cells and hijacking their ribosomes, enzymes, and energy. When a virus enters a cell, it sheds its capsid coat, bares its genes, replicates, and induces the cell's own translation machinery to manufacture more viral protein from the viral nucleic acid. The viral genes and capsid protein self-assemble to form virions, turn the host cell into a viral factory, and strain the cell to the bursting point. The host cell bursts and releases hundreds of virus progeny by the lytic cycle. Some viruses are also capable of lysogeny, and their genomes become integrated into the host cell chromosome. Thus, viruses only need to do two things: they need to have a mechanism for reproduction within host cells, and they need a way to get out of their target cells [76]. Viruses develop a simple way of creating new viruses that require only a minimal investment of molecular machinery.

Viral polymerases play a central role in the viral genome replication and transcription. Several steps in the virus life cycle require the activity of a polymerase. Based on the genome type and specific needs of particular virus, a variety of enzymes are contributed by viral hallmark genes encoding proteins. These are RNA-dependent RNA polymerase (RdRp), RNA-dependent DNA polymerases (RdDp), DNA-dependent RNA polymerases (DdRp), and DNA-dependent

DNA polymerase (DdDp). Viral polymerases slowly transformed RNA viruses to DNA viruses step by step during a recurrent infection. The evolutionary networks of primordial viruses, their recurrent infections of protocells, and their polymerases accelerated the origin of the first cells. Two capsid proteins that are most widely distributed among viruses are the jelly-roll capsid proteins (JRC) and the superfamily 3 helicase (S3H) [75].

6.1 Viruses and Evolution

Viruses are not self-sustaining and need to enter a cell in order to complete their life cycle. Viruses emphasize parasitic roles far more than cooperation in the evolutionary process. Therefore, we tend to regard viruses only as pathogens and thereby dismiss their crucial importance for the evolution of life. Viruses were not only the probable precursors of the first cells, but they have helped to shape and build genomes of all species including humans. The impact of viruses on life is dramatic. The symbiotic relationship of viruses and cells is not always restricted to parasitism, but extends to a wide range of mutualism. The majority of known viruses are in fact persistent and inapparent, not pathogenic (toxic). Many viruses are beneficial to their hosts, providing essential functions in others. Viruses are major drivers of evolutionary transitions [73,75].

The history of life is a story of coevolution of viruses and their cellular hosts. All cellular life harbors diverse genetic parasites including transposons, plasmids, viruses and other selfish elements. The parasite-host coevolution is a major aspect of the evolution of life [75]. The coevolution is often described as an incessant arms race. The billion-year war between viruses and cells is the major source of evolutionary novelties. Viruses evolve, the host adapts, proteins

change, and viruses evade them. It never ends. Many novelties first selected in the viral world might have been transferred to cells as a consequence of the continuous flow of viral genes into cellular genomes. The war has driven a dramatic diversification of viruses and of the host defense system. Viruses have a remarkable capacity to invade, replicate, and evolve within living cells. In response, cells developed an array of defense systems. Viruses and protocells were intertwined since the protein/RNA world. Viral reproduction within a living cell always produces changes in the host cell, sometimes resulting in cell death and sometimes slowly killing the infected cells.

The creative role of viruses in the origin and evolution of life has been known for a century. Viruses are truly nature's genomic laboratory and they help accelerate evolution of the host in a fast lane. Felix D'Herelle, the discoverer of bacteriophages and one of the founders of virology, proposed as early as 1922 that phages or bacteriophages might be the evolutionary precursors of cells [77]. Similarly, J.B.S. Haldane in his 1928 classic *The Origin of Life* suggested an early 'viral' stage of evolution as an integral part of the proposed scenario for the emergence of life from the prebiotic soup [78]. In our discussion of the origin of life, we have followed Haldane's insight.

Viruses are potentially aggressive, selfish-elements, a property that conferred on a parasitic partnership of viruses and protocells; this association has powerful evolutionary potential. Viruses are adept at transferring genetic information between themselves and hosts. Viruses contributed several key enzymes to the host protocells including RNA-dependent RNA polymerase, DNA-dependent RNA polymerase, and DNA-depended DNA polymerase during RNA-DNA transition. RNA viruses might have coexisted with early protocells that still had

RNA genomes. Because viruses have such ancient roots, they preserve a remarkable range of biochemical tricks.

The great billion-year war between viruses and cells are the major source of evolutionary novelties [70]. The reverse transcriptase paved the way to generate DNA; they still generate DNA from RNA in retroviruses, cancer cells, and HIV. Viruses donated DNA and their replicating genes to protocells [75,79]; they might have played a central role in the emergence of eukaryotes and their nuclei [80,81]; they might have been the cause of the partitioning of biological organisms into three domains of life by horizontal gene transfer [82,83]. Their role in information transfer between extant prokaryotes by horizontal gene transfer complicates efforts to build evolutionary trees depicting early life on Earth and to unravel the origin of particular metabolic pathways.

Many viruses have their own, ancient evolutionary history, dating to the protein/RNA world. They are the relics of the protein/RNA world. Viruses possess genes, replicate, evolve, and adapt to particular hosts, biotic habitats, and ecological niches. From prebiotic protocells to unicellular life to human populations, viruses affect life's outcomes and give an ever-changing shape to the fitness landscape, often determining which organisms will survive [73]. Since the beginnings of life, viruses have been the major drivers of macroevolution in all branches of life by horizontal gene transfer across three cellular domains—Bacteria, Archaea, and eukaryotes [82,83]. They comprise the principal source of novel genes in the biosphere [73,84]. Some viruses (e.g., retroviruses) integrate their genetic material into the cell they infect, and if this happens to be a germ line, the viral genome (or its relict) can be maintained essentially forever. About 8% of human genetic material originated from RNA viruses rather than from our vertebrate ancestors [85]. Similarly, retroviruses facilitate the rapid evolution of the mammalian

placenta [86]. Virus-host interaction is an important evolutionary force and played a crucial role in the origin and evolution of life.

6.2 The Prebiotic Origin of RNA Viruses

The origin of viruses is shrouded in mystery, but recent advances in genomics shed light on their ancient ancestry. Viruses have never been detected in fossils, probably because they are too small and too fragile for fossilization. Therefore, the evolutionary history of viruses is difficult to reconstruct. For many years, the central debating point in discussions of the origin of viruses is whether they are ancient, first appearing before the Last Universal Common Ancestor (LUCA), or evolved more recently, such that their ancestry lies with genes that 'escaped' from the genomes of their cellular host organisms and subsequently evolved through independent reproduction [87].

While the geological record cannot offer any clue of when and how viruses originated, genetics provides increasingly strong support for an ancient primordial origin of viruses. First, the vast majority of viruses do not encode genes for ribosomal proteins or genetic evidence of relicts of such genes. Second, this same vast majority of viruses do not contain genetic evidence of ever having encoded enzymes involved in energy metabolism. Third, viral capsid proteins typically do not have obvious homologs among contemporary cellular proteins. Several genes coding for key proteins involved in viral infection as well as major capsid proteins of icosahedral virions are shared by many groups of viruses but are missing in cellular life forms. This is why most viral proteins have no cellular homologues or only distantly related ones. All these combined evidences argue that viruses did not evolve from free-living cells, but arose

independently in the prebiotic world before the first cells. The existence of hallmark genes seems to falsify both the cell degeneration and the escaped genes concepts of viral infection [73,75,88-92]. This primordial origin theory is supported by the strongly inverse relationship between genome size mutation rate across all replications systems, such that pre-Luca genomes were probably both small and highly error-prone and hence RNA-virus like [87].

In the ancient viral world, the flow of virus-specific genes has gone uninterrupted from the precellular stage of life's evolution to the present day. In our view, three major classes of viruses originated in the prebiotic world: (1) positive-stranded mRNA viruses in the primordial protein/RNA world, (2) retroid viruses in the RNA–DNA transition world, and (3) DNA viruses in the DNA world. These ancient viruses emerged in a hydrothermal vent environment in which the mixing and matching of diverse genetic elements was more extensive than it is in any modern biological community. Phylogenetic analyses have shown that that the RNA polymerase, DNA polymerase, and DNA helicase that transcribe and replicate DNA in modern cells were recruited from the viral world [79,88,93].

The idea that viruses are very ancient and have co-evolved with the protocells has recently led to several hypotheses stating that viruses have played a major role in several critical evolutionary transitions [70-73]. Viruses ubiquitously infect all members of the three cellular domains of life, strongly suggesting that protocells with RNA genomes were already the victims of a viral attack. Moreover, viruses infecting cells from the three domains of life—Bacteria, Archaea, and Eukaryotes—suggesting that viruses emerged very early in the prebiotic world before the first cells [71,87,88].

The notion of viral antiquity seems easier to accept for mRNA viruses in the hydrothermal vent environment. High temperatures in the vent environment favored high

diversity of virus-like particles [73]. The only organisms with RNA-coded genomes today are RNA-based viruses, which may shed helpful insight into the protein/RNA world. The virus world retained a distinct flow of genes from the repeated infection of protocells containing RNAs, ribosomes, and proteins. Viruses have maintained their identities and unique parasitic lifestyle ever since, notwithstanding the transfer of many genes between viral and cellular genomes. Several genes that are central to viral replication and structures are shared by all viruses but absent from cellular genomes. In this scenario, the principal lineages of viruses emerged from the primordial pool of primitive genetic elements with a distinct suite of viral genes that retained their identity throughout the entire history of life. Viruses enhanced gene mixing in the prebiotic world by infecting protocells containing RNA and protein molecules, and their subsequent endogenization of them. Therefore, the viral evolution in the prebiotic world is closely intertwined with the origin and early evolution of the cell [90,91,94].

6.3 Pre-virus World

Viruses can be viewed as mobile genetic elements (MGE). The principal lineages of viruses and related selfish agents emerged from the primordial genetic pool of primitive genetic elements in the hydrothermal vent environment, the ancestors of both cellular and genetic elements. The primordial gene pool was crucible for the major virus lineages, where mixing and matching of diverse genetic elements was extensive. Viruses reflect their origin from capsidless selfish replicons, such as plasmids, transposons, and viroids. In this scenario, viruses are direct descendants of primordial genetic elements [71]. These selfish replicons, called viral hallmark genes, encoded capsid proteins with key roles in genome replication, expression, and

encapsidation. These selfish replicons are shared by a broad variety of viruses but are missing from cellular genomes suggesting a flow of virus-specific genes that went uninterrupted from the protein/RNA world of biogenesis to this day. These viral genes are genuine viral hallmarks and can originate either through modification of existing genes or *de novo*. Eventually, diverse protein-coding RNA elements would develop a capsid coat to give rise to the first viruses [71,91].

We concur with the ancient virus world hypothesis [7]. In our model, ancestral viral particles, called here 'pre-virus' could have emerged only when RNA-directed proteins were available in the prebiotic soup in the hydrothermal crater vent environment. In our previous paper, we have discussed how different components of translation machine—tRNA/aaRS/mRNA/ribosome— as well as genetic code emerged step by step inside protocells to initiate protein synthesis [42]. As more and more proteins were manufactured inside protocells, these densely packed biomolecules would exert an osmotic pressure on the phospholipid membrane, occasionally resulting in the rupture of the protocells. The biomolecules would then be dispersed in the prebiotic soup, making it an ideal Nature's genomic laboratory. Different kinds of genetic innovations took place in the genetic pool of the hydrothermal vent that mixed, matched, and evolved new, increasingly complex gene ensembles.

The two important capsid proteins are the jelly-roll capsid proteins (JRC) and the superfamily 3 helicase (S3H), the former is more widespread [71]. We speculate that primordial mRNA gene encoding capsid protein JRC was present in the vent environment (Fig. 8B). New JRC protein genes can originate either through modification of existing genes or *de novo*. Viruses contain many *de novo* genes, namely those in which an existing gene has been 'overprinted' by a new open reading frame; mutations of the mRNA gene led to the expression

of a second reading frame, overlapping the first. Overlapping genes are very common in viral genomes [94].

The origin and evolution of viruses might have occurred on the mineral substrate of the crater floor, where RNAs and proteins were accumulated side by side (Fig. 2). During this prebiotic genomic experiment, naked and fragile mRNA genes might have capped occasionally by proteins that offered protection for stability and durability in the vent environment [7]. This protein coat of mRNA molecules was prelude to the evolution of viral structure. Perhaps, in this milieu of different kinds of mRNAs in the prebiotic soup, JRC capsid genes originated *de novo* within genomes of nonviral mRNAs by overprinting. These capsid genes were capped by proteins on the mineral substrate, transforming them into ancestral viral particles (Fig. 9A). These ancestral pre-viruses had some survival advantages over naked viral genes in the vent environment because of the protective protein coat. Initially, protein coats were random and were not encoded by the enclosed mRNA genes. Perhaps the ancestral pre-viral genes, when engulfed by protocells, could translate it into custom-made capsid protein using ribosomes of the host (Fig. 8B). Primordial viruses could have evolved by encapsidation of these viral genomes.

<Figure 9 about here>

Some of the ancestral mRNA viruses were accidentally ingested by the protocells by infolding of their membranes while searching for food in the vent environment. The engulfed viruses shed their capsid coat, which might be used by protocells for protein storage. This might have been the beginning of endocytosis. The viral genes, on the other hand began to exploit the translation machinery of the protocell to make the custom-made capsid protein.

6.4 The Reproduction Cycle of mRNA Viruses

The emergence of ancestral virus with capsid-coding sequence of proteins was a big evolutionary step and was mediated by ribosome-coding protocells. Here we propose a model for parallel evolution of protocells and the emerging viruses. At the protein/RNA world, the protocells developed different parts of the translation machinery such as mRNAs, tRNAs, ribosomes, several synthetases, and other components for protein synthesis [42]. As more and more proteins were synthesized, various kinds of protocells dominated in the hydrothermal vent environment. Some of these protocells were densely packed with diverse populations of genetic elements, including self-replicating mRNAs, various protein-coding mRNAs, and translation machinery.

A typical viral genome encompasses two core modules that consist, respectively, of genes encoding proteins required for genome replication and proteins involved in capsid formation. In the ancestral virus, the core module might have included all the genes for capsid formation. Initially, protocells with the full set of translation machinery engulfed some of the viral capsid genes from the mineral substrate and inadvertently helped translate their genomes into viral proteins. The engulfed mRNA gene multiplied using replicating enzyme (RNA-dependent RNA polymerase or RdRp) and began to exploit ribosomes of the protocell for synthesis of capsid proteins. Some of the newly created capsid protein strands began to wrap around mRNA as a protective coat where genome could be maintained as a stable structure. Encapsidation was the hallmark of virus's survival without encapsulation. The association of capsids with capsid genomes was a complex process, but it must result in an energetically stable structure. This is the beginning of the viral world, which evolved in parallel with the protocells.

Once some mRNA molecules developed the capsid coat, the first mRNA viruses originated. The capsid affords protection of the viral gene and allows viral genes to gain access to appropriate host cells. As more and more mRNA viruses were created, they exerted osmotic pressure on the phospholipid membrane causing burst of the protocells (Fig. 10A). Slowly, these newly released viruses learned by trial and error how to infect protocells and swap genes with them without inventing their own translation machinery. This innovative short-cut strategy for virus reproduction inside protocells worked efficiently, hijacking the translation machinery of the host. Viruses preferred this parasitic existence from the beginning of biosynthesis and it has continued to proliferate throughout the geologic ages.

Viruses can reproduce only within host protocells, exploiting their ribosomes. The parental virus (virion) gives to numerous progenies, usually genetically and structurally identical to the parent virus. The new generation of viruses began to infect other protocells with an ensemble of translation machinery for their reproduction so that viral mRNAs could be translated into capsid proteins inside the host protocells. Soon, they became obligate intracellular parasites; that is, they developed and reproduced only within the protocells of hosts. They highjacked the host's own machinery to manufacture hundreds of copies of themselves. The reproductive successes of viruses make them archenemies of living cells. Viruses became the capsid-coding particles that began to coevolve with ribosome-encoding protocells. Because of their high mutation rate, viruses were evolutionary accelerators.

We speculate that primordial viruses were single-stranded mRNA viruses that could function both as a genome and as a messenger RNA. It could be directly translated into capsid protein in the host cell by host ribosomes. Like living counterparts, the genome contains relatively few genes, usually between three and ten, including as an RNA-dependent RNA polymerase (RdRp)

or RNA replicase, a viral protein enzyme that synthesizes mRNA from mRNA template. RdRP is an essential protein-encoded in the genomes of all RNA-containing viruses prior to the DNA stage. Pre-viruses would donate an RdRp gene to the host cell making replication of its mRNA much easier than the cumbersome base-pairing method.

We speculate that primordial virions initially didn't kill or lyse their host protocells, but utilized their translation machinery for reproduction of genomes. Some of the likely stages of the cycle of infection were (Fig. 10B):

<Figure 10 about here>

1. Attachment. Capsids on the surface of the mRNA virus attach to the surface of the host protocell (Fig. 10B, cycle 1).

2. Entry via endocytosis. The virus enters the interior of the host protocell through the process of endocytosis.

3. Uncoating of capsid. Inside the protocell, the viral genome emerged from the protein capsid; the capsid in turn destroyed the host mRNA so that viral mRNA occupied its place (Fig. 10B, cycle 2, inset).

4. mRNA copying and protein synthesis. RdRp enzymes copy the viral genome. Energy and ribosome from the host protocell are used to build viral proteins. (Fig. 10B, cycle 2 and cycle 3).

5. Assembly of viral progeny. The viral particles assemble by encapsidation to form progeny virions (Fig. 10B, cycle 4).

6. **Release via exocytosis**. The virus exists in the host protocell by exocytosis (Fig. 9B, cycle 4, inset).

The progeny virions began to infect other protocells to begin the next cycle of infection (Fig. 10B, cycle 5). The life cycle of most viruses is designed to maximize the production of progeny virus particles. Often, the burden of producing a large number of virus particles causes the infected cells to die, the lysis of the host cell. In early stages, primordial viruses probably established a long-term association with the protocell, in which the protocell released a steady stream of viral particles over an extended period of time, benefiting both host and parasite in symbiosis. These ancient RNA viruses had a high mutation rate and underwent evolution and natural selection, just like cellular life, and most of them evolved rapidly. When two viruses infected a protocell at the same time, they might swap genetic material to make new 'mixed' viruses with unique properties. The viral infection of protocell is a prelude to a modern bacteriophage that infects and replicates within Bacteria and Archaea.

This is the beginning of mRNA viruses and their spread in the vent environment. Today, mRNA viruses amount to a large fraction of known viruses including many pathogens, such as the hepatitis C virus, West Nile Virus, dengue virus, and SARS and MERS coronaviruses, as well as less clinically serious pathogens, such as rhinoviruses that cause the common cold [96].

<Figure 10 about here>

6.5 Retroviruses

A retrovirus is a highly derived enveloped particle in which the capsid core contains two identical single-stranded RNA molecules, each RNA carries its genetic and structural blueprint. The virion is 80-100 nm in diameter, and its lipid envelope incorporates and displays the viral phospholipids. The hallmark of a retrovirus is its replicative strategy in the sense that it can

reverse-transcribe its RNA into DNA using its own reverse transcriptase enzyme. This catalyzed transcription is the reverse flow of information of central dogma, hence the name *reverse transcriptase* and *retrovirus*. The new DNA is then integrated into the host cell genome by an integrase enzyme. The host cell treats the viral DNA as part of its own genome, transcribing and translating the viral genes, producing the proteins required to assemble new copies of a retrovirus.

The main virion components of a retrovirus are: a phospholipid membrane and two identical double-stranded mRNA encased in a capsid shell. Each mRNA is typically made up of three genes: the group specific antigen gene (*gag*), the polymerase gene (*pol*), and the envelope gene (*env*). The *pol* gene encodes the three viral enzymes: protease, reverse transcriptase, and integrase—that catalyze the steps of retroviral infection. In retrovirus, multiple protein products are synthesized from a single mRNA species by frame shifting. In between 5′ and 3′ ends of RNA is the protein coding domain, which includes *gag*, *pol*, and *env* encoding regions. These three genes are linked to one through recoding by frameshifting. The ability to make two or more proteins from same mRNA is useful, linking structural (e.g., retroviral *gag*) and catalytic polypeptide (retroviral *pol*). This ribosomal frameshifting mechanism makes retroviral genome more compact (Fig. 11A). Retroviruses have evolved to exploit this translational plasticity in order to regulate their own expressions [97].

<Figure 11 about here>

Once a retrovirus is inside a protocell (a process mediated by protease), it takes over the host's genetic transcription machinery to construct a DNA *provirus*. This process, the conversion of retroviral RNA to proviral DNA is catalyzed by reverse transcriptase and is necessary for proviral DNA into host DNA—a step initiated by the integrase enzyme. Retroviruses can be

pathogens of many different hosts, including humans. A notable retrovirus is human HIV virus, responsible for AIDS.

6.6 Origin of Retroviruses

The known virosphere consists of three principal viral types: the RNA viruses, retroid viruses, and DNA viruses. Horizontal gene transfer (HGT) is rampant among viruses within each of these principal types, but is generally confined to closely related viruses, or viruses (and plasmids) with similar replication mechanisms [98]. There are many examples of mixing and matching in the virus world, but somehow, they so far have been confined to the same type of nucleic acid. A novel virus genome discovered in an extreme hot spring environment suggests recombination of two unrelated groups— between a ssDNA virus and an RNA virus—a natural chimera not seen before [99]. In this hybrid genome, alongside the RNA-derived gene, it contained a gene for DNA replication typical of a DNA virus. Surprisingly, these hybrid viruses are present not just in the acidic lake, but more widespread in a couple of oceanic samples. This find proves that modern viruses can combine information in the two normally separate genetic molecules. And it lends support to the idea that it was viruses that performed the upgrade from RNA and effectively gave rise to DNA. These authors suggest that the hybrid virus may have formed when an mRNA virus, retrovirus, and DNA virus all infected a cell at the same time. The retrovirus used its reverse transcriptase enzymes to make a DNA copy of an RNA virus gene, which combined with the DNA virus's genome to yield this hybrid. In our view, this hybrid virus provides a first glimpse at the ancient viral birth of DNA in the hydrothermal vent environment by the mixing and matching of mRNA virus, pre-retrovirus, and DNA virus.

Retroviruses infect a wide range of animals from fish to humans, and can occasionally leave genomic fossils within their host genome, known as endogenous retroviruses (ERVs). ERVs consist of the genetic material of extinct, or 'fossil' viruses. Our bodies are littered with the shards of retroviruses. Eight percent of our genome is composed of broken and disabled retroviruses, which, millions of years ago, managed to embed themselves in the DNA of our ancestors. Because they no longer seem to serve a purpose or cause harm, these remains have often been referred as 'junk DNA' [73]. Recent phylogenetic study of ERVs placed the time of their most recent common ancestor in the Early Paleozoic [100]. The origin of retroviruses in the Devonian presents an important framework to investigate evolutionary transitions that led to the emergence of the retroviruses. Since vertebrates originated during the Cambrian evolution, ERVs in vertebrate hosts represent the upper limit of the retroviral origin. Molecular precursors of retrovirus probably began in the prebiotic environment billions of years ago in the RNA-DNA Retro world [75]. Retrovirus-like entities are older than the first cells.

Retroviruses bear much similarity to capsidless selfish genetic elements, such as plasmids and various types transposons because they have close evolutionary connections, both share hallmark genes. These hallmark genes encode key components of the viral replication apparatus (such as polymerases and helicases). These retroelements—capsidless genetic parasites—are key to understanding the origin of viral genes [91]. We speculate that these retroelements were self-assembled to a viral gene such as a positive-strand mRNA virus step by step. Positive-sense RNA are particularly suitable for reverse genetics because their genomes are typically infectious in protocells and can be immediately translated by the host's translation machinery. In our model, ancestral retroviruses could emerge only when various retroelements and protein enzymes were available inside protocells.

One of the critical enzymes synthesized inside protocells was the reverse transcriptase (RT) enzyme. RT is an RNA-dependent DNA polymerase. Due to the limitation of the genome size that can be packaged in the virus shell, viral polymerases are generally active as a single protein capable of carrying out multiple functions related to the viral genome synthesis. Here we propose a simple evolutionary scenario for the origin of retroviruses. The vast class of retroelements is united by a single conserved gene, the RT gene, which also defines the key feature of their reproduction cycle, reverse transcription [91].

The ancestral stage of a retrovirus is called here ‘pre-retrovirus,’ a retroid, non-enveloped mRNA virus with a minimum functional design. In our model, pre-retrovirus was derived from mRNA virus inside an infected protocell. Perhaps, viral RT gene evolved *de novo* inside protocell [95]. Most likely, the mRNA viral gene and the viral RT gene merged by accident into one single-stranded mRNA and was enclosed in a capsid shell. In this fused viral gene, one gene was used for synthesis of structural capsid protein, the other for the RT gene for synthesis of reverse transcriptase enzyme (Fig. 11B, C). The RdRp enzyme could be used by pre-retroviruses to replicate their genomes.

The ability to make structural protein and enzyme from the same mRNA gene had distinctly selective advantages over two separate genes performing similar functions in a double-stranded RNA virus. Most likely, protocells, by that time, had developed several weapons to ward off viral attacks by destroying viral genomes. In response, the pre-retrovirus altered their RNAs in such a way as to thwart attacks from the protocells. Moreover, this linking of two genes, capsid gene and RT gene, into one mRNA strand allowed a novel compact, space-saving mechanism of the genome. This ability of linking two genes into one was achieved by the ribosomal frameshifting mechanism as discussed earlier [97]. A pre-retrovirus had now become

a single-stranded mRNA virus with a capsid coat to play a formidable parasite during protocell infection (Fig. 11B, C). With the emergence of retroviruses, the protein/RNA world transformed itself into the 'retro world' [91].

<Figure 11 about here>

6.7 The Origin of DNA Viruses from mRNA Viruses

DNA can be considered modified RNA because there are only two chemical differences between RNA and DNA molecules. The first is the removal of a single oxygen atom from RNA (ribonucleic acid) to generate deoxyribonucleic acid, or DNA. The second difference is the addition of a methyl (CH_3) group to the nucleotide base uracil (U) to generate thymine (T).

Forterre [79,88] suggests three stages in the evolution of DNA from RNA through viral infection: the RNA world, the RNA-to-DNA transition, and the DNA world. In the RNA world, RNA viruses emerged from the RNA cell. In this model, LUCA is considered as the primitive RNA cell from which the RNA virus emerged. The RNA virus gave rise to three lineages of the DNA viruses in the RNA-DNA transition. These three lineages of DNA virus evolved in parallel into three domains of life—Bacteria, Archaea, and Eukaryotes.

Forterre [79,88,101] proposes that DNA viruses evolved directly from RNA viruses in two steps during the RNA-to-DNA transition. The critical enzymes for RNA to DNA conversion was supplied by the viral world. In the first step of the RNA-to-DNA transition, the deoxyribose in DNA was generated from the ribose in RNA by enzymes called ribonucleotide reductases, which converted RNA to U-DNA in the genome. In the second step in the evolution of DNA was the conversion of the uracil base to thymine by thymidylate synthases, forming T-DNA (DNA

containing thymidine). The emergence of thymidylate synthase activity in some U-DNA virus lineages produced viruses with the modern form of T-DNA. As these new strains of T-DNA viruses infected protocells, the host genomes gradually transformed from U-DNA to T-DNA. Once deoxyribonucleoside triphosphates were available, their assembly into DNA-like chains followed fairly rapidly. The idea that both ribonucleotide reductases and thymidylate synthases were first encoded in viral genomes and were later transferred to protocells is compatible with phylogenetic analyses of these enzyme families. Thus, viruses donated both DNA genomes and their replicating genes to the host protocells [79,88]. The hypothesis of a viral origin for DNA could explain why many DNA viruses encode their own ribonucleotide reductase and/or thymidylate synthase [102]. According to this model, the RNA world and LUCA appeared at the same time with the existence of a few homologous DNA informational proteins in the three cellular domains. Cellular DNA and its replication machineries originated in the DNA world via transfers from DNA viruses to RNA cells. Three such independent transfers led to the origin of Bacteria, Archaea, and Eukaryotes.

Though Koonin et al. [75], accept the viral origin of DNA and criticize the non-cellular LUCA concept of Forterre in the RNA world. Instead, they favor the traditional view that LUCA emerged after the first cells from which all three domains descended (Woese 1998). Moreover, they emphasize the crucial role of retroid RNA in the origin of DNA. DNA could not have emerged from the RNA world without reverse transcription. The retro elements must have been among the first classes primordial viruses that evolved in the primordial genetic pool in the protein/RNA world after the advent of translation, when several kinds of enzymes were synthesized. Integration of such elements into host genomes must have coevolved with the increase in the size of the DNA genome. We find the retroviral infection model of the origin of

DNA [75,91] very attractive and plausible. In our view, retroid RNA such as pre-retrovirus might have invented DNA step by step by converting RNA with the help of the RT enzyme. We suggest three stages in the viral evolution, combining both Forterre's and Koonin's model that gave rise to DNA from RNA: the protein/RNA world, the Retro world, and the DNA world (Fig. 12). In our view, the existence of RNA-only cells (protocells) of Forterre (2006) seem to meet formidable difficulties. A more parsimonious scenario includes the peptide/RNA world [42].

<Figure 12 about here>

In the protein/RNA world, viral genes might have evolved *de novo* in the vent environment from pre-existing mRNA. These mRNA genes encoded JRC capsid protein for a protective shell for stability and durability to form the first virus inside protocells (Fig. 9). These ancient mRNA viruses began to infect protocells to increase their own populations in the gene pool (Fig. 10).

In the Retro world, pre-retroviruses developed the ability to stitch two genes into one. The first gene encoded the JRC capsid protein, but the second gene encoded the reverse transcriptase (RT) enzyme. The RT enzyme could generate complementary DNA from an mRNA template. The pre-retrovirus would play the crucial role for conversion of RNA to DNA during recurrent protocell infection, and subsequently incorporated into the host genome as a provirus. The reverse transcriptase enzyme, also called RNA-directed DNA polymerase, that catalyzed the conversion of RNA to DNA was not present in the host protocells, but was delivered by the pre-retroviruses that converted RNA of the host genome directly into DNA. The RT enzyme performed three sequential biochemical activities: RNA-dependent DNA polymerase activity, ribonuclease H (RNase H), and DNA-dependent DNA polymerase activity. Collectively these activities enabled the RT enzyme to convert single-stranded RNA (ssRNA) into single-

stranded DNA (ssDNA). In our model, pre-retroviruses gradually modified its mRNA to DNA during recurrent retroviral infection. As soon as a pre-retrovirus invaded a protocell by endocytosis, its capsid coat was dissolved. The capsid gene segment would be translated into capsid protein and the RT gene segment into the RT enzyme, thus exploiting the ribosome of the host. The RT enzyme would then transcript its own single-stranded RNA template to single-stranded DNA (ssDNA); the RT would make another strand of cDNA from another mRNA; two strands of cDNA then combined to make a double-stranded DNA (ssDNA). The conversion of ssDNA to double-stranded DNA was mediated by DNA polymerase (DdDp) (Fig. 13). The information contained in a retroviral gene is thus used to generate the corresponding protein via the sequence:

mRNA—> DNA —>mRNA—> protein.

In the DNA world, once the dsDNA virus appeared, the role of the RNA pre-retrovirus was gradually replaced by DNA viruses. DNA viruses possessed a full set of independent DNA replication enzymes: a helicase to unwind the DNA helix, two DNA polymerases (DNA Pols) to replicate the two strands, and a primase to form RNA primers that DNA Pols extended. As dsDNA viruses infected an RNA protocell, its DNA genome was replicated into multiple copies. Each DNA genome then replaced the mRNA of the protocell, transcribed to the new generation of mRNA that translated into protein. This is the beginning of the DNA world, when DNA replaced mRNA as the major genome of the protocell. The DNA protocell followed a conventional pathway of flow information as in a cell: DNA—> mRNA —> protein.

<Figure 13 about here>

The abundance of genetic parasites along with the presence of defense systems in all cellular life forms strongly suggest their coevolution might have started in protocellular stage.

The DNA viruses might have emerged as a novel survival strategy. When the RNA protocells were confronted with an invading pre-retrovirus, they might have protected themselves by a number of defense mechanisms and developed an ancestral immune system. Immunity of viral infection allowed the RNA protocells to proliferate. In response, pre-retroviruses might have invented DNA to ward off attacks from the hosts. For pre-retroviruses, DNA might have offered a very powerful, immediate benefit: That pre-retroviruses might have immediate fitness benefit for substituting DNA replacing its RNA genome [103].

It thus appears that the transition from RNA to DNA genomes occurred in the viral world, and that protocellular DNA and its replication machineries originated via transfers from DNA viruses to RNA protocells [101,103]. The DNA virus living in a carrier state in an RNA protocell probably lost the genes for capsid proteins and became established as DNA plasmids. These plasmids were later transferred to RNA protocell and incrementally transformed into DNA by recurrent infection. In such a scenario, the RNA protocell was transformed, from within, into the DNA protocell cell [88]. With DNA genome and its transcripted mRNA and ribosome, protocells began to function as a DNA protocell that began to synthesize protein. The coded genetic information began to flow from DNA to RNA to protein, beginning the classical central dogma of molecular biology.

There must have been a long transition period during which DNA in a protocell, helped by natural selection, progressively took over the replicative and storage functions of RNA. The DNA genomes are usually far larger than those of the RNA that allowed more storage capacity for biological information. Once deoxyribonucleotide triphosphates (dNTPs) were available, it is likely that their assembly into DNA-like chains would have followed fairly rapidly. DNA has

intrinsically higher replication fidelity, which allows genomes to increase in size and therefore complexity [63].

7. The Advent of DNA

DNA was derived from RNA. In fact, these two molecules are so similar that transformation of RNA to DNA requires two constituents of RNA to be replaced by two close relatives: ribose by deoxyribose, and uracil (U) by thymine (T). As to the information, RNA could be transferred to DNA by reverse transcription, as happens in cells infected with retroviruses. We have already discussed the origin of the DNA virus from the RNA protocell by pre-retroviruses. Although the origin of DNA from RNA via retrovirus is widely accepted [71,88,91], there are some dissenters. The latest twist in the origin of DNA debate is that RNA and DNA might have appeared together in the prebiotic world from building blocks without the assistance of viruses.

Powner et al. [102] proposed a novel pathway for the prebiotic synthesis of several microcomponents in the assembly of DNA molecules from a mixture of the chemicals thought to have been present in the sulfur-rich prebiotic environment. In this environment both RNAs and DNAs emerged simultaneously, not one after another. They argued that switching of RNA nucleotides into DNA nucleotides needed special enzymes that were costly to produce in terms of energy and material. On the other hand, if DNA molecules were present alongside RNA molecules, this problem of fundamental switching from conversion of RNA to DNA could be solved. The possibility of abiotic route for the synthesis of deoxyribonucleic acid, according to these authors, provides a new perspective for the origin of DNA.

A similar view has been expressed by other authors. For example, Xu et al. [104] suggested *de novo* assembly of DNA from the building blocks of life in the prebiotic environment. They identified a compound presumably present in the prebiotic Earth called thiouridine that could have linked DNA nucleotides into chain-like DNA. Prebiotic phosphorylation of the 2-thiouridime molecule gave rise to nucleotides of DNA via photoreduction. In their view, both RNA and DNA may have arisen all at once in first life forms.

The possibility that DNA and RNA might have evolved concurrently appears to be the less parsimonious explanation than DNA following RNA. The DNA must have appeared later than the RNA molecule, because RNA degrades and mutates easily [39]. The backbone of single-stranded RNA is much less stable than the equivalent structure in double-stranded DNA. Its enhanced stability and longer molecular sequence give DNA greater fidelity and increased memory in its information storage system. RNA replication is intrinsically error-prone compared with DNA replication. DNA was selected over RNA based on its expanded capacity to store information and its dramatically improved error rate during replication [49].

7.1 The Evolutionary Advantage of DNA over RNA

As the genetic diversity increased and the translation perfected, DNA protocells must have faced dilemmas: how to increase the information storage capacity of mRNA, and how to separate translation from replication? Unlike RNA, DNA is a poor catalyst. RNA acted as both template and catalyst, whereas DNA can function only as a template. There was only one way out: division of labor between the RNA catalyst and the DNA template. Information storage and replication become the prerogative of DNA, while the utilization of this information for protein

synthesis and other functions remains the province of RNA. This is when DNA took charge in protocellular function, when replication of information is entirely dissociated from its expression. Because genes no longer had to serve messengers, they could be joined together in strings of increasing length to maximize storage capacity of information [63,105]. RNA is more versatile, acting in both the storage and replication of genetic information. However, during this transition from RNA to DNA, the DNA lost its catalytic power but transformed into a stable, long-term storage molecule for genetic data. The functional separation between replication and catalysis must have been a tremendous improvement in DNA molecules and favored by strong selective forces. The division of labor between template and catalyst is a fundamental attribute of all living systems. This fundamental property of life is believed to have been absent in the protein/RNA world. Another advantage of DNA is that it allows a selective expression of individual genes by way of transcription, keeping the DNA molecule intact. In contrast, during translation, each mRNA molecule is destroyed.

With the advent of DNA, all the genes could now be kept as stable, double-stranded threads, which are held together by complementary base pairing and are twisted in a double helix. The extremely stable structure doesn't allow DNA to mutate rapidly unlike RNA, thus converting to an efficient information storage structure. Intact stretches of DNA have been recovered from fossil bones that are at least 700,000 years old. The DNA acts as a permanent record—a blueprint containing the information needed to build protein and run the cell. A segment of DNA is transcribed to short-lived mRNA for translation responsibility, but DNA preserved the information storage system intact. Storage in mRNA without the possibility of retrieval would have been useless. Hence the need for transcription. The stored information

could thereby be recovered from mRNA for translation. Transcription is entirely dissociated from translation.

The DNA is completely protected, but mRNA strands are continually made, broken down, and recycled. Deoxyribose in its sugar-phosphate backbone makes chains of DNA chemically more stable than chains of RNA, so that much greater lengths of DNA can be maintained containing multiple information of genes without breakage. The DNA double helix replaced RNA as a more stable molecule for storing the increased amounts of genetic information required by protocells. These distinctions enable the two molecules to work together and fulfill their essential roles during protein synthesis.

The capacity and durability of single-stranded RNA molecules were severely limited, with a high rate of error during its replication. A backup copy of RNA was required to protect the original code or restore the prototype if damaged. Because DNA is a double-stranded molecule, it is much more stable than RNA, and the presence of two strands provides a way of repairing genes. A damaged copy of the gene is restored using the complementary strand, the second copy of the gene, as a guide. This need for a repair mechanism may have been the selection pressure driving the creation of the double strands in DNA, conferring both stability and immortality.

The RNA is much more susceptible to chemical transformation. The DNA backbone is less prone to hydrolysis because it lacks the nucleophilic 2-hydroxyl group, so it represents a more chemically stable material and is less prone to mutation. As a result, DNA replicates with greater intrinsic fidelity than RNA, thus allowing more information storage. RNA polymerase, the enzyme that generates RNA from a DNA template, has no proofreading activity, and whereas DNA has many repair mechanisms, RNA has none. DNA soon superseded RNA as the carrier of genes and became the dominant bearer of genetic information. Being much longer than RNA,

DNA can store information on thousands of genes. Once DNA was established, it allowed genes to become longer and more complex. The fidelity of DNA replication is orders of magnitude greater than that of RNA replication. Unlike the single-stranded RNA molecule, in the DNA molecule, two antiparallel strands that are complementary in their nucleotide sequences are paired in a right-handed double helix, with about 10 nucleotide pairs per helical turn.

Once DNA molecules appeared on the scene in the DNA world, they took on the role of the primary genetic molecules, superseding the RNA molecules, which became intermediaries between DNA and proteins. These new DNA-containing protocells rapidly diversified into large populations that easily outcompeted the RNA-based protocells [49,106].

7.2 DNA structure

The double helix structure of DNA, proposed by Watson and Crick [107], is an iconic image based on two paired DNA strands that are complementary in their nucleotide sequence. The immortal coil of two intertwined strands creates a molecule with the shape of a twisted ladder. The sides of the ladder are made of the sugar-phosphate backbones of two strands. The four nitrogen bases are adenine (A), thymine (T), guanine (G) and cytosine. They project off of the sugar-phosphate backbone and join together by hydrogen bonds, forming the ‘rungs’ of the ladder (Fig.14A). Each strand of DNA is a long sequence of four bases. The order of these bases is what determines DNA’s instructions, or genetic code. Both strands of double-stranded DNA store the same biological information. The two chains are held together by hydrogen bonds. A pairs with T by forming two hydrogen bonds; and G pairs with C by forming three hydrogen bonds. Because the AT and GC pairs are equal length and fit identically into the double helix

(like rungs on a ladder), the diameter of the helix is uniform, 2.0 nm (Fig. 14B). These two strands are complementary, but they go in opposite or antiparallel direction. The DNA sequence directly encodes each protein's structure, which determines its activity. The spiral is 'right-handed'—twisting upward as if driven by a right-handed screw, a chiral feature. All these features increase DNA's stability and thus its effectiveness as a reliable information-bearing molecule [106].

Solving the structure of DNA immediately revealed how the two fundamental processes of inheritance and mutation worked at a molecular level [106,107]. That is, a DNA sequence could be faithfully copied and passed on because of the base present at each position on one strand determined its complement on the other strand. Mutations result from errors in copying processes, where the wrong base or extra bases{(s)}{necessary?} gets inserted, or a bases{(s)}{necessary?} may be deleted, generating a change in the DNA sequence.

Two important functions evolved with advent of DNA: transcription and replication. Transcription is the first step in gene expression. It involves copying a gene of the DNA sequence to make an mRNA molecule for protein synthesis. Replication, on the other hand, is the process by which the double-stranded DNA molecule is copied to produce two identical DNA molecules. As discussed earlier, the critical enzymes needed for transcription and replication were provided by viruses [88].

7.3 DNA Transcription

The storage of information in DNA without the possibility of retrieval would have been useless. The RNA can function as both template-directed polymerase and template, whereas the

DNA can function only as template. As their genetic diversity and sophistication increased, protocells needed a division of labor between the DNA template and its catalyst mRNA to function properly. In other words, replication had to be separated from translation [105]. Therefore, a transcription mechanism emerged. The information stored in DNA must be recoverable in mRNA for translation. Transcription is the first step of DNA based protein synthesis, in which a particular segment of the DNA template is copied into a complementary mRNA by the DdRp enzyme (RNA polymerase). This is the transfer of genetic information from a segment of the DNA, called a gene into mRNA. Transcription proceeds in 5´ to 3´ direction. That transcript, an mRNA gene is then used to produce a protein.

The first step in the process is the unwinding and separation of the two strands of the DNA helix. RNA polymerase then travels along the length of the strand of DNA and binds complementary RNA nucleotides to it, until a complete strand of mRNA is formed, encoding at least one gene (Fig. 14C). Once a gene is transcribed, the newly made mRNA is dissociated from the DNA template. As the mRNA molecule is formed, the DNA helix zips itself back together. This new generation of mRNA is then translated by ribosome to synthesize a protein chain. The DNA transcription is the beginning of the process that ultimately leads to the translation of the genetic code via mRNA. This is when the central dogma emerged combining the two-step process, transcription and translation, by which the information flows from DNA to mRNA to protein [39] (Fig. 10D). In our previous paper, we have discussed in detail the origin of the translation and the genetic code [42].

<Figure 14 about here>

Although DNA generated new proteins, this production was balanced by the loss of proteins through their degradation. The interaction between proteins and DNA was of mutual benefit and was important for the survival of the protocell.

7.4 DNA Replication

DNA replication is basically a polymerization process that requires a template and a primer (DNA polymerase), and the product of the reaction is a new strand of DNA that is complementary to the template strand. DNA replication is central to the reproduction of all cellular life forms and many viruses. Every time a cell divides, the DNA polymerases are required to help duplicate the cell's DNA, so that a copy of the original DNA molecules can be passed to each daughter cell. In this way, genetic information is passed from generation to generation. Cellular DNA replication systems are broadly classified into two types, one in Bacteria and the other in Archaea/Eukaryota. In contrast, double-stranded DNA viruses have a much broader diversity of DNA replication systems. Both Forterre [79] and Koonin [89] suggest that the protocells acquired the DNA replication core enzymes such as helicase, primase, and DNA polymerases (Pols) from the DNA viruses during their recurrent infection of protocells.

The double helix structure of DNA had striking implications for the processes of DNA replication. There are some similarities between transcription and replication. In both cases, the DNA double helix is untwisted when the hydrogen bonds between the bases are broken. However, there are major differences between these two processes. Transcription copies the DNA into the mRNA, whereas replication makes another copy of the DNA. In transcription, strand separation is mediated by the RNA polymerase, whereas in translation, the DNA

polymerase takes this role. Both processes involve the generation of a new molecule of nucleic acid, either mRNA or DNA. However, the function of each process is very different, with one involved in gene expression and the other in cell division. Transcription was the precursor to replication. DNA replication must have evolved as a prelude to cell division.

The DNA replication process is well known in literature (Freeman 2005) and its basic idea will be discussed briefly to highlight its role in the first cell division and the origin of life. The antiparallel structure of DNA is important in DNA replication because it replicates the leading strand in one way and the lagging strand the other way. The DNA replication is performed by the replisome, a complex molecular machine composed of numerous enzymes. It is a complex process that requires the coordinated effort of a team of enzymes for unwinding, separation, replication, and rewinding the helix. The replisome is composed of two replicative polymerase complexes, one of which synthesizes the leading strand, while the other synthesizes the lagging strand. DNA replication requires at a minimum, a helicase to unwind the DNA duplex, two DNA polymerases (Pols) to replicate the two DNA strands, and primase to form RNA primers that DNA Pols extends. Replication occurs in three major steps: the opening of the double helix and separation of the DNA strands, the priming of the template strand, and the assembly of the new DNA segment.

During separation, the two strands of DNA double helix uncoiled at a specific location called the origin. The initiation of the DNA replication occurs when an initiator enzyme called a helicase unwinds a short stretch of the DNA double helix using the energy of ATP hydrolysis that breaks apart the hydrogen bonds between the bases of the DNA strands and opens up like a zipper in one direction at the Y junction, or replication fork, to form two strands: a continuous leading strand and a small, discontinuous lagging strand. In DNA, these two strands are

antiparallel, the 3’ end of the leading strand is paired with the 5´end of the lagging strand. Each strand serves as a template for a new strand. (Fig. 15).

In the next step, several enzymes and proteins work together to prepare, or prime the strands for duplication. A doughnut-shaped DNA polymerase requires a primer—a starter strand of RNA to which they can add new nucleotides. An enzyme called primase synthesizes a short stretch of RNA that acts as a primer for the DNA polymerase; this RNA primer can simply match ribonucleotides directly by complementary base pairing on single-stranded DNA.

Synthesis of the leading strand is straightforward after an RNA primer is in place. A DNA helicase, powered by ATP hydrolysis, propels itself rapidly along the leading template DNA strand forcing the DNA to open the DNA helix ahead of the replication fork. The helicase moves into the replication fork, which unzips ahead of it by another enzyme called topoisomerase that relieves the twisting forces. As the replication fork moves, the DNA polymerase can move continuously along this arm of the Y in 5{’} —> 3´ chemical direction, adding complementary nucleotides to the 3´ end of that strand to produce a new daughter DNA molecule. As the replication continues, it creates two double-stranded helices, which are an exact copy of each other.

<Figure 15 about here>

The synthesis of the lagging strand is more complicated, because the DNA strand here is antiparallel, so that the DNA polymerase must work in the opposite of the replication fork, 3´—> 5´ chemical direction. The synthesis of the lagging strand starts when a primase synthesizes a short stretch of RNA that acts as a primer. Synthesis requires repeated priming and extension of the lagging strand discontinuously as a series of Okazaki fragments (Fig. 15B). The DNA polymerase then adds bases to the 3´ end of the lagging strand. Here the synthesis is

discontinuous, where the DNA polymerase can synthesize relatively small, discontinuous stretches in 3´—> 5´ direction. These short stretches of new DNA are called Okazaki fragments, which are then linked by ligase enzyme to a continuous whole. Once all the bases are matched up, an enzyme called Ribonuclease H strips away the RNA primers. The gaps where the primers were are then filled by yet more complementary nucleotides. Finally, an enzyme called DNA ligase seals up the sequences of the DNA into two continuous double strands.

The result of the DNA replication is two DNA molecules consisting of one new and one old chain of nucleotides. This is why DNA replication is described as semi-conservative, half the chain is part of the original molecule, half is brand new. Following replication, the new DNA is automatically coiled tightly around an axis to form a double helix for stability and greater compactness. In bacterium, circular DNA is a higher-order helix-upon-a-helix, known as a superhelix.

The DNA replication apparatus has evolved over billions of years through ‘trial and error’ since its inception in the protocell in the protein/RNA world. The complex process by which DNA is replicated today could not have been the original version of replication, so a simpler, more primitive mechanism remains to be discovered. One possibility is that in the DNA genome of the protocell was a very small to a bare minimum, few hundreds of nucleotides where replication was less cumbersome, requiring few unwinding and rewinding enzymes and DNA polymerases for replication of each strand (Fig. 15C). It required two replisomes for bidirectional replication. These short nucleobases of two DNA strands were uncoiled and completely separated into two strands, each strand can serve as a template for making a new complementary strand. The net result was formation of two new double double-stranded DNA sequences that

were exact copies of the original double-stranded DNA. In this way, the double-helical DNA could be copied precisely.

As more and more nucleobases were added to the DNA with time to make longer and longer strands, one enzyme after another was selected and added to the mix of primordial enzymes to the replication fork, presumably for fine tuning to increase the speed, control or overall accuracy of the replication process.

8. Biological Stage

8.1 What is Life?

Life, as we know it today, is an outcome of interactions between genetic opportunities, metabolic capabilities, and environmental changes. How to define 'life' is a sweeping question that affects whole branches of biology, biochemistry, genetics, and ultimately, the search for life elsewhere in the Universe. Yet, there is currently no consensus regarding the definition of life. We need a working definition of life so that we can identify the first cells from their protocellular ancestors. In 1944, the Nobel physicist Erwin Schrödinger, prior to the discovery of DNA, defined life as that which 'avoids the decay into equilibrium.' [108]. This definition refers to the *Second Law of thermodynamics*, which says that entropy always increases. But living things, said Schrödinger, are able to postpone this trend by taking its nutrients by metabolism, an open system, unlike the closed system of the physical world. Life is an open system that exchanges matter and energy with the surrounding environment. However, attempts to define life with

thermodynamics have so far failed to distinguish clearly between the living and the non-living. A fire also fits this definition, but it is not considered 'living' [109].

Schrödinger argued correctly that organisms must run a sort of computer program, which is what we now call the genome. He speculated that genes were some kind of 'aperiodic solid' that contained some version of an 'elaborate code-script' that specified all of the future development of the organism. Von Neumann [110] following this argument of genetic software, found striking similarity between software/hardware of a cell and those of a computer, and created 'Universal Constructor' to mimic the cell division. Today, all living cells utilize DNA as the replicable repository of genetic information, express the information by transcription of the DNA to the mRNA, and translate the mRNA into proteins by the same metabolisms by the same mechanisms, including the same genetic code [63]. Thus, the genetic code was the 'elaborate code-script' conjectured by Schrödinger.

Information is the currency of life, where information flows from DNA to RNA to proteins. Living systems store and process the information and transmit it to their offspring. Information is the language of life. Recently, it has been argued that the genetic software provides the singular definition regarding what life is [111]. Biological information separates life from non-life. In this view, life emerged in that instant when information gained control over the biomolecules. It has become clear in recent years that the biological world is computational at its core. Algorithms, or instruction sets to synthesize proteins, are found in every cell. Digital storage of the molecular information system is the key to defining life and understanding its origin. The key mechanism is the origin of the genetic code. The algorithms of life are a much more complex system than today's most sophisticated computer. The information system emerged about four billion years ago during prebiotic synthesis and is still working perfectly [42].

It is universal in all life forms from bacteria to humans. The information-directed protein synthesis is a unique signature of life. Although it is difficult to define what makes life so distinctive, there is a general agreement that its informational aspect is a key property, perhaps the key property.

But there are other attributes of life too. All life has a cell membrane, nucleic acids for replication, and proteins for metabolism. Life maintains itself by making more of itself. Life reproduces life. Life's continuous production of itself, called *autopoiesis*, is a unique hallmark [112]. DNA is an unquestionably important molecule for life, but the molecule itself is not alive. When a DNA molecule produces another DNA molecule, it is replication. When a cell divides into identical daughter cells, we call it reproduction [48]. This is why viruses are not alive. They are not autopoietic. Too small to self-maintain, viruses do not metabolize. Viruses do nothing until they {entire}{enter?} an autopoietic entity: a bacterial cell, the cell of an animal or plant. They lack sufficient genes and translation machine to maintain themselves. Thus, the definition of life is not straightforward.

A living system integrates three critical functionalities in an interactive system: (1) the *cell* membrane maintains an identity over time by localizing all its components and protecting them from the environment; (2) the *protein* uses free energy from its environmental resources in order to maintain itself and grow by metabolism, (3) and the *DNA* carries inheritable information and controls cell division. In order to reproduce successfully, a cell must be able to copy and transmit all of its genetic information to each of its daughter cells. The DNA replication allows cells to do this. During cell division, the replication of the DNA molecule makes occasional errors (mutations). The altered base sequences then produce the variations in a population of

cells that are essential for evolution by natural selection. The ability to evolve is the final aspect of any definition of life.

Although we have an intuitive understanding of what it means for something to be alive, it is difficult to come up with a precise definition of life. Mindful of the centrality of evolution, NASA defined life as 'A self-sustaining chemical system capable of Darwinian evolution.' This minimalist definition captures the essence of life. One of the most distinctive characteristics of this definition of life is the concept of Darwinian evolution to establish a new level of a biological system. NASA's definition has a universal application which is not limited to Earth where RNA/DNA/protein-based life emerged. While this definition is open enough to include a wide range of potential life forms anywhere in the Universe, it also makes hard to design a simple test for life or identifying a fossil. If we find a microfossil in the Archean hydrothermal chert, or in the crater basin of Mars, can we use this definition to ascertain whether this fossil is the primordial life? Probably not. We need some morphological attributes also. Life is what is common to all living beings, which are made of cells. Here, we have expanded NASA's definition of life based on its attributes: 'Life is a self-sustaining, DNA-based information system enclosed in a plasma cell membrane, which is capable of reproduction and Darwinian evolution.' In case of the microfossil mentioned above, if it retains a distinctive cell membrane and a distinctive shape (spherical, rod-shaped, or spiral), our arguments for identifying primordial life become more robust.

There is no bounded entity smaller than a cell that is capable of independent reproduction. Thus, the RNA cell, often depicted as the first cell [79], does not fit into our definition of life. We regard the RNA cell as a protocell, a long way from the first cell. The lonely RNA cell, without the assistance of peptides, could not create the genetic code, the soul of the biological

information system [42,45]. Life might have originated from interactions between RNA and peptides. Although replicating molecules, like nucleic acids, and catalytic molecules, like enzymes, are essential for life, by themselves they are not alive. Life is when these components are highly organized and contain specialized coordinated components into an interactive system, then it takes on the properties of life. In our view, the first cell was a highly organized living system using DNA and RNA as genetic material, 20 genetically encoded amino-acids, ribosome for template-directed protein synthesis, and membranes that allowed for chemiosmotic coupling.

Many of the problems in defining life boil down to the fact that we have only one example—life on Earth. The Earth is the only planet currently known to support life. We know what is life from a single sample from our planet, which is surprisingly quite uniform. As revealed by its remarkable biochemical and morphological similarities, life on Earth has a common origin. All Earth-based organisms use cell membranes that separate the interior of a cell from the outside environment, nucleic acids for hereditary information, proteins to control biochemical reaction and catalytic activity, and identical phosphorous-containing molecules (such as ATP) to store energy. It's the same biochemistry from bacterium to human [109]. If life is detected on Mars or other celestial bodies inside our Solar System, perhaps our definition of life would be refined or modified.

8.2 First Cells

The origin of modern cells is arguably the most challenging and important problem in the field of Molecular Biology. Cellular life depends on a number of fundamental properties that must have been achieved step by step in protocells. The ability to divide must have been a

property of protocells from the moment of their existence. However, the mechanism of cell division with the DNA replication creating identical daughter cells is a breakthrough innovation that defines the first life.

The fossil record suggests that the first cells, a signal of a prebiotic evolutionary event appeared on Earth almost four billion years ago. Morphologically these Archean microfossils look like bacteria [113]. The central question about the origin of life has been how the first cells arose from the DNA protocells and how these first cells began to reproduce, proliferate, and evolve. The birth of the living cell was the culmination of the prebiotic synthesis. It was the turning point in the history of Earth, transforming a rocky planet into *Gaia*—a living planet. The ability to multiply in exact copies of itself is what makes life so completely different from anything else in the known Universe. As discussed earlier, the prebiotic chemical evolution from cosmic ingredients to a single-celled living organism probably took place in the hydrothermal crater vent environment. These pioneer heat-loving microbes – hyperthermophiles – thrived around the volcanic vent of the crater that belched out several toxic gases such as CH_4, NH_3, SO_2 and H_2S, providing chemicals for nutrients and warmth. These hyperthermophiles in the vent environment were probably the first life forms on this planet [4].

Here we reconstruct the first cells from the combined evidence of the prebiotic evolutionary history, the Archean microfossils, and the morphology and function of modern bacteria. From the fossil record of the oldest microbes, it appears that the first cells were probably simple spheres or rods. They were likely a single-celled microbe with minimalist design, but they acquired the essential components of cellular structure for integration and reproduction. The plasma membrane was a simple, double-layered cell wall, and lacked extensive, complex, external and internal membrane systems. The outer membrane was leaky. It

was filled with porin proteins, which formed small holes through the membrane allowing passive diffusion. These holes were big enough for nutrients and ions to pass through by passive diffusion, but small enough to keep the machinery of the cell inside. The membrane encased the cytoplasm in which all soluble components were packed inside. Like bacteria, the first cell might have contained only one circular chromosome in the cytoplasm – a large DNA molecule, packed closely in the nucleoid. Most likely, this chromosome was longer than the diameter of the cell in which it was encapsulated, so the circular configuration of the chromosome was a space-saving device for the crowded mass of DNA. To fit into the nucleoid, the DNA double helix coiled on itself with the aid of enzymes to form the highly compact 'superhelix' circular structure, which was enclosed tightly in a condensed area of nucleoid (Fig. 16). Outside the chromosome, the most prominent structure in the soluble portion in the cytoplasm was dispersed ribosomes; other smaller molecules included various enzymes and other translation machinery for protein synthesis. Some molecules were devoted to producing energy from environments. Perhaps the flagella were not yet invented as the fossil record suggests; the first cell lacked this propelling device to power movement.

<Figure 16 about here>

It was known for many years before the detailed structure of DNA was determined that DNA is replicated during cellular reproduction. It is important here to draw a sharp distinction between replication and reproduction. Molecules can replicate, but only cells can reproduce [48]. The first and most important attribute of life is reproduction. If something is alive (autopoietic), it can make a copy of itself. Reproduction with variation is an essential characteristic that distinguishes life from nonlife. The reproduction of the first cells was the watershed event in the history of life. Without reproduction, life would quickly come to an end, so reproduction is

essential for the continuity of life. The biological stage was the tipping point, the boundary between the information age and the biological evolution. How do nonliving biopolymers such as membranes, nucleic acids, and proteins morph into living organisms—organizing into cells, reproducing, then growing, and evolving? The first life was integrated molecular systems of interconnected parts to produce homeostasis and discovered ways to respond to stresses, both physical and chemical. The major components of the first life were a plasma membrane enclosing the core enzymes and information carrier genomes forming the integrated systems that included metabolizers and the ability to capture energy from vent environments. The first life was autopoietic.

Building new proteins was one of the main tasks of the first cells. Over half of the molecules inside the *Escherichia coli* cell are involved in one way or another with the synthesis of proteins [114]. The making of various types of proteins was one of the most important events for a cell because protein not only forms structural components of the cell. It also composes enzymes that catalyze the production of the remaining biomolecules necessary for life. In the early cell, different genes were active that produced only those proteins that were needed in the cell. All proteins in the first cell were encoded in one large circular DNA. The cell must control when and where each gene was used. The information held in the DNA genome was highly regulated, Perhaps, a host of repressors and activators interacted with each gene, determining when it would be used to create protein.

The first life orchestrated how to control gene expression at any step between the synthesis of the mRNA and the activation of the gene product. Gene expression can be controlled at any step between the synthesis of the mRNA and the activation of the first gene product. In the first cell, the prebiotic information systems became increasingly sophisticated to process more

and more advanced levels of biological information. In the peptide/RNA world, information flows from the mRNA to the protein [42]. With the emergence of DNA, the central dogma is established; information flows from DNA to mRNA to protein. With the emergence of the first cell, information flows from cell to DNA to mRNA to protein.

Four steps occurred during the flow of information in the first cell for the gene expression, represented by arrows in the following expression:

Cell —> DNA—> mRNA—> protein —> activated protein.

The arrow from cell to DNA represents the selection of a gene by the cell to make a particular protein. The arrow from DNA to mRNA represents the transcription of that gene to mRNA. The arrow from mRNA to protein represents translation, in which ribosomes read information in the mRNA and use that information to synthesize a protein. The arrow from protein to activated protein represents post-translational modifications.

Life came to a climax when algorithms of cells took control. It began in the peptide/RNA world with the origin of the genetic code, then refined with DNA replication, and culminated with the emergence of the first cells. It is the informational role that is the key to transforming nonliving protocells into life. A key feature of biological information in algorithmic life was feedback from the environment. Changes in gene expression allowed first cells to respond to environmental changes. Gene expression could be controlled at three levels: transcription, translation, or post-translation (post-activation).

The emergence of the first cells was the defining moment in the history of our planet that made Earth unique—the *Blue Marble* in the Solar System. Once the first cells formed and began to reproduce, Darwinian evolution began to diversify early life into different species that invaded different environments. The evolution of reproduction in the first cells was the turning point in

the early history of life, heralding the biological stage and ultimately leading to biodiversity. Life began to move from the crater basin to the global ocean, diversified and proliferated [4].

8.3 Cell Division

Life is organized into cells that grow and divide. The advent of cell division into two identical cells with a complete set of genetic machinery defines the emergence of the first cells from their protocell precursors. Although protocells could divide by physical forces, it was not controlled by DNA replication. The chromosome replication was tightly linked to the first cell division. We can infer the origin of the primordial cell reproduction from the modern bacterial fission and protocell division.

To reproduce, bacterial cells divide into two genetically identical cells to propagate by binary fission. A common bacterium, *Escherichia coli*, which colonizes the mammalian intestine, is an excellent exemplar of the cell cycle and binary fission. The cell cycle of a bacterium is the series of events that takes place in a cell that lead to the duplication of its DNA, the segregation of copied DNA, the splitting of the parent cell's cytoplasm, and its division into two identical daughter cells. The genome consists of a single, circular DNA chromosome that must be replicated and then allocated into the daughter cells; the cytoplasmic contents must also be divided to give both new cells the machinery to sustain life. *Escherichia coli* can divide every 20 minutes.

Most bacterial cells spend their time cycling between a state of calm (interphase) and the dividing phase, in what is known as the 'cell cycle'. The starting point of replication, the origin, is close to the binding site of the chromosome to the plasma membrane (Fig. 17). A cell must

grow to twice its original size during interphase, and this growth is coordinated with the duplication of the chromosome. In bacterial fission, the circular DNA chromosome is attached to the cell wall, near the midpoint of the cell. DNA replication occurs bidirectionally as the cell grows and elongates. Before binary fission occurs, the cell must replicate its chromosome and segregate these copies to opposite ends of the plasma membrane. The cell division and the DNA replication were coordinated in such a way that the distribution of new DNA copies to each daughter cell was ensured. The seemingly simple process of bacterial growth and division requires an impressive orchestration of functions by several regulatory proteins. These core cycle regulators can cause key events, such as DNA replication or chromosome separation, to take place. They also make sure that the cell cycle events take place in the right order.

Bacterial cell division depends on the formation of a furrow in the plasma membrane with infolding of membrane and cell wall. A constriction develops by the Z-ring between the duplicated chromosomes so that one is included in each of the two daughter cells. The fact that the bacterial chromosome is anchored to the plasma membrane makes this separation possible without the construction of an elaborate mitotic apparatus. This contractile Z-ring orchestrates the separation of the chromosomes and the cell division in the bacterium. A septum is formed between the nucleoids, extending gradually from the periphery to the center of the cell. When the new cell walls are in place, the daughter cells separate, completing the cell division [106,115].

<Figure 17 about here>

Laboratory simulations hint at a solution to the primitive cell division mechanism. Lipid vesicles extracted from the Murchison meteorite undergo spontaneous primitive cell division in the laboratory, with no external forces acting upon them [15]. When a mixture of these cosmic vesicles, amino acids, and nucleic acids was shaken, the vesicles trapped the organic molecules

inside them and began to interact. Therefore, these vesicles can apparently take in substances from outside themselves through their lipid walls, and use them to build new walls and new contents. A large vesicle mimics a primitive kind of cell division.

The physics of 'chemically active' droplets, which cycles chemicals in and out of the surrounding fluid may shed light on the origin of protocell division [116]. The team studied a theoretical model for behavior of a liquid droplet in a chemically disequilibrium system. This 'active droplet' behavior differs from passive and more familiar tendencies of oil droplets, which join together into bigger and bigger droplets without dividing. On the other hand, these chemically active droplets can grow to a stable size by taking resources from the environment. Droplet growth eventually leads to instabilities – linked to the changing shapes of the droplets. The droplet keeps elongating and pinches in at the middle, which has low surface areas. Eventually, surface tension causes it to spilt into a pair of droplets. This process of dividing droplets somewhat mimics the spontaneous vesicle division from the Murchison meteorite [15].

In a laboratory simulation, a genome-rich vesicle increased in size at the expense of an empty vesicle. When its greater size imposed too much osmotic stress, pearling instability developed, and the stretched vesicle divided into two, each daughter vesicle retaining some of the original genomic contents [115]. Recent work on model protocell membranes has demonstrated that vesicles can grow as filamentous structures and divide spontaneously under mild shear forces, and that with photochemical stimulation, a robust 'pearling' mechanism produces many small daughter vesicles [61]. Self-replicating membranes can divide either spontaneously or under the influence of external environmental forces [51], and high environmental shear forces can cause vesicles to divide.

Synthetic biologists use simple ‘protocells’ to study the origin of cell division, but previous models were not able to reproduce both the genome and the membrane sustainably. Kurihara et al. [117] proposed a recursive self-proliferating model protocell that represents a step towards eventual production of model protocells that are able to mimic cell division. They used a novel system by fusing the self-reproducing vesicles with feeder vesicles, thus allowing the vesicle composition to be sustained over multiple generations. Because of competition, the larger vesicle grows more quickly and fuses with the feeder vesicles. Thus, feeding the protocells by vesicle fusion offers a practical pathway for indefinite self-reproduction [118].

The mechanism of cell division is a complicated multistep process. A new study shows that a mutant bacterium can reproduce without a wall or division machinery, supporting the idea that primordial cells could have divided using physical mechanisms, such as simple shearing alone. Leaver et al. [119] generated a mutant strain of *Bacillus subtillis* that lacked cell walls. Although the Z-ring retains in this mutant form, it does not participate in cell division. They found that these cells divide without the Z-ring but by an extrusion-resolution mechanism. This novel form of cell division provides insights into how early forms of cellular life may have proliferated. This pattern is strikingly similar to ‘pearling instability’ seen in lipid vesicles. The study supports a model in which the constriction of the Z-rings is dependent in wall synthesis.

Although the structure of the bacterial cell is very different from that of the DNA protocell because of billions of years of evolution, we may assume some basic common mechanism of protocell division from bacterial fission. The protocell might have achieved a rudimentary form of cell division using physical mechanisms alone, and was modified with the availability of proteins. Initially, the DNA protocell division may have been an accidental burst, when the cell size increased as the cell accommodated more and more complex biopolymers,

until it reached an unstable size. The surplus molecules generated inside the protocell, causing it to bulge, and the protocell continued to elongate and became pinched in the middle. Eventually, the surface tension caused the cell to split into a pair of daughter cells. As the cell divided, the two new cells were not necessarily identical, because there was no mechanism in place to ensure an equal distribution of the parent cell's contents. The dissimilarity in the daughter cells was an advantage at this stage of evolution because it promoted a diversity upon which natural selection could operate, but true cells required mechanisms that guaranteed identical daughter cells [120].

DNA replication was the major driver of symmetrical binary fission in early cells. However. it took many generations of DNA cells to invent the cell reproduction by countless trial and error. We can speculate how the DNA cell might have learned how to reproduce into an identical daughter cell. Perhaps, initially the cell grew in size, then elongated for cell instability, and pulled apart by physical forces, but divided asymmetrically without DNA replication, so one daughter cell might have intact DNA, while the other was devoid of it (Fig.18). In the second generation of the attempt, the DNA might have replicated, but it was not coordinated with cell division, so one daughter cell might have two chromosomes, the other got only cytoplasm without any DNA. Eventually, the DNA protocell learned how to coordinate DNA replication with cell division, probably aided by few replicators. As DNA was replicated, the replicated DNA molecules were segregated at the two ends of the cell in an energy-dependent process, forcing the cell to grow, elongate, and pull apart, making the equatorial region highly stretched, weak, and constricted. DNA replication exerted osmotic pressure on the cell wall by increasing the concentrations of entrapped biomolecules, and DNA separation actively dragged the nucleotides apart. At that time cytoplasm was enriched with FtsZ protein that aided to constrict the membrane at the midpoint. The Z-ring established the location of the division site, acted as

the scaffold for the division apparatus, and provided the contractile force to precisely orchestrate the binary fission. As the contractile Z-ring closed like a purse string, the protocell cytoplasm was divided into two, completing the cell division (Fig. 18). For the division to produce viable daughter cells, it must be coordinated in time and space with other major events of the cell cycle, such as DNA replication and segregation.

<Figure 18 about here>

The first cell was self-sustaining, DNA-based and chemically sophisticated, possessing many of the housekeeping proteins, and capable of mutation and Darwinian evolution. It had developed capacities for harnessing energy from the hydrothermal environment. However, these primordial organisms were continuously infected by viruses and developed some immune systems for their survival. The coevolution of viruses and the first cells were the source of innovation, gene enrichment, and diversity of the early life. The first cells were stabilized, perfected in cell division, multiplied, mutated for innumerable generations, and spread across hydrothermal systems in the young Earth.

The first cells became the earliest self-sustaining organisms and their emergence was the turning point in the early history of life. Suddenly, the first cells began to dance and reproduce and multiply quickly, colliding with each other and crowding the vent environment (Fig. 19). Reproduction was an essential life process for the first cells because it allowed them not only to survive but to continue as a population. Reproduction of the first life was the most momentous event in the early history of our planet, transferring the barren alien rock into a living world. Since then, life itself has in many ways helped to shape our planet. Earth is much more complex than all other Solar System objects that we know of because of life. As soon as life acquired the ability to spread and multiply, it also carried the potential to change the environmental chemistry

from which it had emerged. The early history of life on Earth may have been characterized by coevolution of microbial metabolism and atmospheric composition. Earth is the only planet we know of that can support life. This is an amazing fact, considering that it is made out of the same matter as other planets in our Solar System, was formed at the same time and through the same processes as every other planet, and gets its energy from the Sun.

<Figure 19 about here>

8.4 Hyperthermophiles

The first traces of life on Earth dates back to the Early Archean, about four billion years ago (Fig. 18A). Although the recognition of ancient microfossils just by morphology turned out to be difficult, the geologic setting of these fossil-bearing strata indicates the hydrothermal environment. Only hyperthermophiles would have been able to thrive and survive in this extremely hot environment [4,21].

The habitats hyperthermophiles (superheat-loving microbes), which are the most primitive living organisms, may shed new light on the oldest ecosystems on our planet—the cradle of life. Discovered in 1977, submarine hydrothermal vents astounded many scientists when it was discovered that the hyperthermophilic bacteria and archaea thrive in these deep, dark, anaerobic, hostile and volcanic environments through their ability to utilize the chemical nutrients that arise from the hot vent fluids interfacing with cooler sea water as a source of energy [58]. Today, hyperthermophiles are found in geothermally heated subterranean rocks such as boiling hot springs of Yellowstone National Park, hydrothermal impact crater-lakes, and submarine hydrothermal vents along the mid-ocean ridge. They grow optimally above 80º C and

exhibit an upper temperature border of growth up to 113º C. Hyperthermophiles have certain heat-stable enzymes (that are proved to be very important in biotechnology) and unusual rigid membranes that are specially geared to working in high temperatures. The cell membrane contains high levels of saturated fatty acids to retain its shape at high temperatures. Based on their growth requirements, hyperthermophiles were probably the most primitive living organisms that could have existed on early Earth about four billion years ago [21].

9. The LUCA: The Genetic Portrait of the Ancestor of All Life

Darwin's theory of common descent with modification is the central pillar of modern evolutionary biology [121]. It states that all life on Earth which has ever lived on Earth has descended from one original primordial form that diverged with time – like tree branches from a single trunk. Darwin recognized that species not only evolve but also divide. As species evolve, they split and diversify through time, increasing morphological divergence like the branching pattern of a tree. He presented a metaphorical tree in his book that showed how species change through time from the common ancestor. Darwin's tree is a visual representation that shows how species are related by a common descent.

Darwin left to later biologists to figure out what the real evolutionary tree looked like. Before the development of sequence-based molecular methods, it was impossible to know the evolutionary relationships connecting all of life and thereby to draw a universal evolutionary tree. Although the topology of the tree of life has changed over time as more and more genetic and proteinic information of organisms are available, they all confirm Darwin's theory of common descent. A genetic portrait of the ancestor of all living things has been slowly emerging in recent

years. This venerable ancestor, the last universal common ancestor or LUCA, was most likely a single-cell, bacterium-like organism that lived in the hydrothermal vent environment.

But how can we formally test the idea of common descent? There is a compelling list of circumstantial evidence – for example the universal genetic code and the universal use of the same biomolecules – DNA, RNA, and protein, and homochirality of certain key molecular structures from bacterium to human. The organisms that exist today contain numerous proteins that are clearly homologous, most notably ATP synthase. The organisms themselves must therefore be homologous, descended from common ancestors, the most recent of which we understand as LUCA. The proposition that all extant life is genetically related—is perhaps the most fundamental premise of modern evolutionary theory, providing a unifying foundation for all of life science.

Darwin's dream of the tree of life was realized on the grandest scale when Woese [119] proposed that all cellular life could be proposed into three separate fundamental groups of domains—the Bacteria, the Archaea, and the Eukarya, based upon sequence comparisons of ribosomal RNA (rRNA) sequences. According to the 'three-domains tree' the Eukarya and Archaea are more closely to each other than they are to Bacteria. The 'three-domains tree' is the most visible image depicting the diversity of cellular life, but it has not gone unchallenged. An alternative 'two-domains tree,' in which Eukaryotes are nested inside Archaea, has gathered support from recent phylogenetic analyses [123].

Molecular phylogenetics have now provided overwhelming evidence that all living organisms on Earth descended from a single ancestral form, the last universal common ancestor (LUCA). The concept of LUCA is central to the study of early evolution of life's origin, yet its nature and phylogenetic placement in the tree of life are controversial. Woese [124] identified

LUCA as ‘progenote,’ an organizational level similar to an RNA protocell that preceded with the first cell. A similar view of LUCA has been suggested by Forterre [88]. Today, LUCA is considered as a sophisticated organism with a complex structure recognizable as a true cell. LUCA is probably a bacterium-like organism that could help establish how early life on Earth got established. LUCA is not thought to be the *first* cell, but the *last* before the divergence of the two-domain tree – Bacteria and Archaea (Fig. 20B). There was life before LUCA, when the first cell was mutating and evolving into diverse populations [125]. LUCA is the most common ancestor of all living organisms, but only one of many early prokaryote cells, which is still extant, whereas the others became extinct.

The question of whether or not all the life on Earth has an ultimate common ancestor is a subtle one, complicated by the phenomenon of horizontal gene transfer (HGT). Present evidence suggests that this is a widespread phenomenon in the bacterial world. Some scientists believe that LUCA was not an organism, but a collection of diverse organisms exchanging their genes by rampant HGT more or less without constraint; LUCA in this view looks like a tangled tree of life contradicting its monophyletic origin [83]. The same objection applies to the possibility that certain genes that were not present in LUCA arose later, in separate lines, by convergent evolution. With due regard to these uncertainties, we have enough information to sketch a portrait of the universal ancestor.

Theobald [126] analyzed the vast array of molecular sequences now available from the three domains of life using powerful statistical techniques and concluded a monophyletic origin of LUCA regardless of HGT or multiple origins of life. He studied amino acid sequences from a set of 23 universally conserved proteins found in the three domains of life. He then applied standard programs for inferring evolutionary trees. His study was based on several simple

assumptions about how diversity of protein arose. He applied a model of selection theory to a molecular phylogeny that favored the existence of a single origin of LUCA over a wide suite of alternative hypotheses. According to Theobald, LUCA was a microbe living in the early Archean world from which all life evolved.

Weiss et al. [20] genetically analyzed 6.1 million protein-coding genes and 286,524 protein-coding clusters from sequenced prokaryotic genomes of various phylogenetic trees, and identified 355 protein families that were probably common to LUCA. The team realized that HGT between bacteria and archaea about four billion years ago masked much of LUCA's original genetic signal. Genes found in both bacteria and archaea could have been shared through HGT and hence would not necessarily have originated in LUCA. The team searched for 'ancient' genes that have exceptionally long lineages but do not seem to have been shared around by HGT on the assumption that these ancient genes should therefore come from LUCA. Once they finished their analysis, they found only 355 genes that definitely belonged to LUCA and can tell us something about how LUCA lived. LUCA shares these genes with two groups of modern microbes: *Clostridia*, a genus of thermophilic Bacteria, and the methanogens, a group of hydrogen-metabolizing Archaea. Most likely, LUCA lived in an anaerobic hydrothermal vent rich in H_2, CO_2, and iron. LUCA was chemosynthetic and autotrophic, deriving free energy in the form of redox potentials and pH gradients from the vents. Furthermore, hyperthermophiles have adapted to vent conditions, so maybe life began here.

Although the physiology and habitat of LUCA in the hydrothermal vent environment suggested by Weiss et al. [20] has been endorsed by previous researchers, their conclusion about the portrait of LUCA – a progenote, only 'half-alive'– has created a great deal of controversy. Such a small number of genes (~355), of course, would not support life as we know it, and critics

immediately jumped onto its apparent gene shortage, pointing out that essential components capable of nucleotide and amino acid biosynthesis, for example, were missing. Gogarten and Deamer [127] have rightly criticized that these authors reintroduced an old misconception view of LUCA, a progenote, which had been rejected by most scientists. LUCA had evolved far beyond the origin of life; it was a full-fledged prokaryote cell that could accomplish the complicated task of synthesizing proteins.

LUCA was a sophisticated organism preceded by a long period of Darwinian evolution. The emergences of the first cells and LUCA are separate events, the former led to the latter. As we discussed earlier, there was life before LUCA. Once LUCA appeared, it quickly gave rise to two domains – Bacteria and Archaea (Fig. 20B). Although LUCA is long gone, its closest relatives may still be with us.

LUCA itself was a hyperthermophilic chemoautotroph, a view supported by molecular phylogeny [21,126-130]. Allowing for uncertainty, LUCA was not very different from modern hyperthermophilic organisms that appeared 4–3.5 Ga, before the split between Bacteria and Archaea (Fig. 20B). The geological setting of the earliest microbial life is generally considered to have been a hot hydrothermal crater vent environment, thus supporting the physiology and habitat of LUCA [4,20,21].

10. Radiation of Archean Microbial communities

The conditions at the surface of the young Earth were suitable for the emergence and evolution of life. The early atmosphere of Eoarchean Earth was dominated by CO_2 and N_2, with smaller amounts of H_2, CO_2, and CH_4; the climate was probably similar to that of more recent

times [2]. The best sources of information to infer the presence of a stable liquid water veneer on early Earth, which was the prerequisite for the origin, evolution, and propagation of life, are sedimentary rocks. Minerals, such as zircons, and water-lain sediments in the ancient Hadean/Archean crust indicate that liquid water was prevalent as early as 4.3 billion years ago [17].

The fossil record of Early Archean life had been erased mostly by plate tectonics, meteorite impacts, weathering, erosion, and recycling of the crust. The older the rock, the lesser is the chance of its preservation of these delicate fossils. The oldest volcano-sedimentary record of Earth's early Archean history provides tantalizing clues to the emergent biosphere [32,131]. The hydrothermal systems of ancient Greenstone belts including the Nuvvuagittuq Craton of Canada, Akilia-Isua Craton of West Greenland, the Pilbara Craton of western Australia, the Kaapvaal Craton of South Africa, and the Singhbhum Craton of India have yielded the earliest evidence of life. This finding is consistent with molecular biology that the universal ancestor and the ancestor of bacteria were hyperthermophiles [21,128-130]. The evolution of Bacteria and Archaea from LUCA illustrates the astonishing durability of prokaryotic life. They are built to grow and multiply as fast as possible.

Our discussion of early Archean life is entirely based on prokaryotes—Bacteria and Archaea. Although prokaryotes are microscopic, they outweigh macroscopic creatures at least ten-fold. They are extremely diverse in their metabolism—the range of nutritional and respiratory strategies, which are adapted later by plants, animals, and fungi. They can swap genetic information without the elaborations of sex in eukaryotes. They are far more variable in appearance than is often imagined: some are spherical, some are rod-shaped, some have whip-

like appendages called flagella for swimming. They can communicate by quorum sensing to determine whether other microbes are close to them.

The hydrothermal crater vent environment was highly favorable for early habitation during the last stage of the bombardment period that enabled early life to get a foothold in a bunker, leading to sequestered communities. Their ecosystem was energized not by solar power, but heat from vents and nourished by hydrogen sulfide, methane and other toxic chemicals spewed from vents. Experimental study indicates evidence for the hyperthermophilicity of ancestral life [128]. Recent phylogenetic analysis suggests that the first enzymes were fully adapted in those extreme hot environments [132]. Modern hydrothermal crater settings represent a snapshot of the primordial niche for early life. These subaerial hydrothermal lakes are widely colonized by hyperthermophilic bacteria and archaea and provide a glimpse of early history of life on our planet. Similarly, the red, orange and blue shades around the geysers and mud holes in Yellowstone National Park disclose hyperthermophiles' ecology on land in extreme harsh environments. The pigments of hyperthermophiles are carotenoids with red, yellow and orange hues that make them colorful.

In a previous paper, Chatterjee [4] reviewed the known Archean biosignatures, microfossils, and stromatolites from hydrothermal deposits during the Eoarchean (4.0–3.6 Ga) and the Paleoarchean (3.6–3.2 Ga) eras in different parts of the world. These microfossils give us valuable clues to the successive evolution of the microbial world (Fig. 20A, B). Three distinct microbial regimes can be reconstructed from the Archean fossil records in stratified sequences of the hydrothermal crater vents from the bottom to the top (Fig. 20C). They harnessed different sources of environmental energy to grow and nourish. In the earliest stage, hyperthermophilic bacteria (such as Thermotogales) and archaea (such as Methanococcales) emerged concurrently

with LUCA and adapted independently a benthic hyperthermophilic life style around the vent of the hydrothermal crater basin. These hyperthermophiles were chemosynthetic and harnessed energy stored in chemicals such as iron, hydrogen, sulfur, and methane, from the vent environment. These simple organisms consumed the chemicals from the vent environment and employed enzymes to speed up chemical reactions, which released energy that the organisms harnessed along with ATP for their metabolism. Hyperthermophilic bacteria employed sulfur reduction metabolism in the vents, whereas archaea produced methane as a metabolic process.

In the second stage of microbial evolution, there was a gradual change of temperature gradient and niche from extremely hot vent site (hyperthermophilic) to the near surface of the crater basin at moderate temperature (thermophilic) when primitive, anoxygenic, photosynthetic bacteria (such as green and purple sulfur bacteria) began to appear in a hydrogen-sulfur world, but tapped infrared from the Sun. These early photosynthetic life forms were chemoautotrophic, which used bacteriophyll pigment to harvest light, and energy from chemical substances from vent environments to build glucose. These ancestral bacteria used hydrogen sulfide, not water, to carry out a type of photosynthesis that did not produce oxygen.

In the third stage of microbial evolution, advanced, oxygenic, photosynthetic bacteria (such as cyanobacteria) emerged at the water surface of the crater-lake that adapted to the normal water temperature niche (mesophilic). As they began to harness solar energy using chlorophyll pigment, they began to spilt water molecules, used hydrogen from water to build glucose and produced oxygen as a byproduct. These phototrophs converted sunlight into biologically useful energy in the form of chemical gradients and reduced molecules. The evolution of aerobic photosynthesis is undoubtedly the single most important metabolic innovation in the history of life. Some anaerobic photoautotrophs, such as halophiles (salt-loving bacteria), are the most

closely related to the cyanobacteria. They live in an extremely salty environment and carry out a peculiar type of anaerobic photosynthesis using rhodopsin as a pigment [113].

Though closely related, the anoxygenic photosynthesis and oxygenic photosynthesis differ in their metabolic pathways. Both combine hydrogen and carbon dioxide to build glucose, but the hydrogen comes from different sources. In the primitive anoxygenic process, hydrogen is supplied by hydrogen gas (H_2), small organic compounds, or hydrogen sulfide (H_2S). Light energy is harvested by bacteriochlorophyll pigments in anoxygenic bacteria to build glucose, the universal cellular fuel.

A more complicated photoautotrophy evolved next, based on chlorophyll-containing light-sensitive photosynthesis [113]. Here, hydrogen is always provided by water, and for that reason oxygen is liberated by breakdown of the water molecule, and hydrogen atoms are split away. A key component in oxygenic photosynthesis is the oxygen-evolving complex that is based on manganese compounds exploiting a transition from Mn_4O_4 to Mn_4O_6 that suggest the vicinity of a hydrothermal vent [133]. Cyanobacteria are the inventors of ‘green plant photosynthesis’ using chlorophyll and the ultimate source of breathable air. Photosynthesis is the end result of a long chain of evolutionary selection at the molecular level that shows gradual evolution of pigment with a changing niche to harness energy.

These vertically stratified temperature regimes and niche partitions in crater basins encouraged different kinds of microbes to adapt thermotolerance and diversity (Fig. 20). As cyanobacterial life invaded the ocean surfaces from a hot to normal temperature regime, the associated enzymes coevolved to keep life’s chemical reactions going and shifted their optimal temperature range gradually to a cooler environment [132].

Both kinds of photosynthetic bacteria, anoxygenic and oxygenic, created distinctive stromatolite horizons in the Pilbara and Kaapvaal sequences that indicate their microbial activity and distinctive habitats. Anoxygenic photosynthesis predates the oxygenic photosynthesis in standard phylogenetic models. Possible microfossils with cyanobacteria-like filamentous morphology and stromatolites are present in the Pilbara Craton of Australia around 3.5 Ga [113]. Today, anoxygenic photosynthetic bacteria and cyanobacteria coexist in layered stromatolitic communities, because they have different light-capturing pigments. Life on present-day Earth is largely dependent on the products of oxygenic photosynthesis. The development of anoxygenic and then oxygenic photosynthesis would have allowed life to escape the hydrothermal setting and invade the global oceans by switching to solar energy from hydrothermal energy.

As anaerobic and aerobic bacteria invaded the shallow-water of the global ocean, they competed side by side to harvest sunlight. However, cyanobacteria produced molecular oxygen, a gas toxic to their anaerobic neighbor, killing countless species of anaerobic bacteria and archaea, because their energy-generating metabolic processes were not coupled with the consumption of oxygen. In fact, the presence of oxygen actually poisoned some of the key enzymes of anaerobes. These anaerobes took refuge in the hydrothermal vents and remained there forever, a survival strategy. The anoxygenic photosynthetic bacteria retreated initially to the deeper part of the water column on the murky ocean floor, shielded by a cover of water where there was no free oxygen. They survived by fleeing the oxygen warfare. Later, they learned how to coexist with oxygenated bacteria by hiding below the cyanobacterial mat, where the infrared could penetrate. Certainly, anaerobes didn’t vanish from the Earth, but they were vanquished to low-oxygen environments. It was perhaps the first of the mass extinctions that life

would face on our planet, and the killing agent was deadly oxygen gas produced by the cyanobacteria.

Some of the earliest evidence for Archean microbial ecosystems stems from putative stromatolites 3.5 to 3.2 Ga, which display similar structures to modern microbial mats. The earliest microbial mats consisting of bacteria and archaea may have formed as biofilms in hydrothermal vents. As photosynthetic bacteria left hydrothermal vents and invaded the shallow ocean, the microbial mats developed profusely in this environment, allowing widespread colonization of the globe and the creation of further aerobic habitat. A key feature of modern-day microbial mats and stromatolites is the fact that they are oriented and stratified in relation to sunlight. Thus, there is reason to believe that ancient stromatolites may have functioned similarly to microbial mats today. Microbial mats are a unique ecological niche representative of early life on Earth. Rapid nutrient cycling across microgradients coupled with putative niche differentiation within mat layers enables diverse metabolic processes to occur in spatial proximity.

<Figure 20 about here>

On a broad scale, the origin of oxygenic photosynthesis by cyanobacteria led to the rise of oxygen on Earth during the early Proterozoic (~2.4-2.3 Ga), played a major role in the oxygenation of ocean and atmosphere, and paved the way for oxygen-breathing microbes such as respiratory bacteria. The *Great Oxidation Event* is when the atmosphere first became oxygenated. However, oxygen level only reached somewhere between 0.2 to 2% by volume, not today's 21% [109]. Respiration is opposite to photosynthesis. Whereas photosynthesis employs energy from the Sun, respiration releases energy. In photosynthesis, cyanobacteria combine carbon dioxide and water to form glucose and oxygen. In respiration, aerobic bacteria consume sugar and

oxygen to gain energy. One biochemical trick that evolved around 2.5 billion years ago to take advantage of oxygen is still being used for respiration by all animals including humans.

The history of life is largely microbial. Microbes once ruled the whole world. The Earth has been a bacterial world for at least the last 3.5 billion years. The diversity of life today is a result of the dynamic interplay between genetic opportunity, metabolic capability, and environmental changes. For most of their existence, Earth's habitable environments have been dominated by microbes and subjected to their metabolism and evolution. Microbes gave life its initial foothold, and they engineered the planet for our use, taking in carbon dioxide and giving off oxygen, day in and day out for billions of years until there was enough oxygen in the atmosphere to support larger life.

11. Discussion and Conclusion

Life depends on the intricate interplay of myriads of different biomolecules, but how such molecular networks arose at the origin of life remains a mystery. In a previous paper, we discussed how molecular symbiosis, a mutually beneficial interaction between different classes of molecules such as peptides, RNAs and lipids could have emerged, leading to the information system, genetic code and translation machine in a series of steps for protein synthesis [42]. The major theme of this paper is the prebiotic synthesis in the *Protein/RNA world* that gave rise to the first primitive cells. A protein/RNA world is widely accepted as a probable step in the early evolution of life. The newly synthesized proteins functioned as enzymes, membrane proteins, and DNA-building proteins. They were the nanobots of the protocells controlling metabolism, transport, communication, and structure of the membrane. Enzymes helped to catalyze and

mediate critical molecular evolutions in the prebiotic synthesis and replaced ribozymes for efficiency. The overwhelming number of RNA-directed protein enzymes in this stage became a driving force in evolution that led to four major innovations in biochemical makeup during biogenesis: (1) the emerging of the phospholipid membrane from the fatty acid membrane and its subsequent conversion into the leaky and efficient plasma membrane; (2) the origin of prebiotic cytoplasm; (3) the beginnings of the virus world; and (4) the advent of DNA.

The first innovation is the origin of the modern cell membrane. Fatty acids of protocell membranes could evolve into diglyceride fats and then phosphatidate acid by serial non-enzymatic reaction. With the availability of enzymes and proteins, the phosphatide acid evolved first into phospholipid membranes, then plasma membranes that allowed selective permeability for certain molecules to cross in and out by diffusion.

The second innovation is the emergence of primitive cytoplasm. Once different kinds of proteins were synthesized by translation machinery and available for prebiotic synthesis, they began to accumulate in the primitive cytoplasm inside protocells. This prebiotic cytoplasm became the storehouse for lipids, nucleic acids, enzymes, proteins, translation machinery, as well as water and salt. The accumulation of various biomolecules made the cytoplasm gel-like substance. Many metabolic reactions, including protein synthesis, and transition from RNA to DNA began to take place in the primitive cytoplasm.

The third innovation is the beginning of the virus world. The evolutionary history of viruses represents a fascinating, but controversial topic. Recently, several investigators proposed virus-first hypothesis before first cells appeared on Earth. I concur with this prevailing view that pre-viruses arose during the prebiotic stage. In my model, newly synthesized proteins helped create a virus world in the vent environment in parallel with the protocells, where naked mRNA

molecules were capped occasionally by a thin protein coat to protect fragile mRNAs from degradation. High temperatures in the vent environment favored high diversity of virus-like particles [73]. Initially, protocells with a full set of translation machinery engulfed some of these ancestral viruses to initiate endocytosis. Inside the protocells, these ancestral viruses shed their capsid coats and their genes began to coax protocellular machinery to create custom-made capsid proteins, while the viral genes replicated by enzymes. These capsids began to coat viral genes, thus creating a new generation of mRNA viruses, which escaped protocells by exocytosis. These new generations of viruses began to infect other protocells, benefiting the genomes of both host and parasite in a relentless war.

The fourth innovation is the rise of DNA from the virus world in a series of steps. From the mRNA virus, pre-retrovirus evolved, where the genome contains not only a capsid gene but an additional gene, a reverse transcriptase (RT) gene merged into one single-stranded mRNA and was encased in a capsid shell [97]. In the fused viral gene, one gene was used for synthesis of structural capsid protein, the other RT gene for synthesis of reverse transcriptase enzyme that could be used for converting the mRNA of a protocell to the DNA protocell. It appears that the transition from RNA to DNA genomes occurred in the *Retro world*, and that protocellular DNA and its replication machineries originated via transfers from DNA viruses to RNA protocells. Recent works suggest that viruses have played a major role in the DNA replication enzymes and possibly of DNA itself [70,79].

The transition from the RNA world to the DNA world was a major event in biogenesis. The invention of DNA required the appearance of enzymatic activities for both synthesis of DNA precursors, retro-transcription of RNA templates, and replication of single and double-stranded DNA molecules. The origin of DNA protocells from the DNA virus paved the way for

the origin of central dogma, where the information began to flow from DNA to RNA to protein. DNA is more resilient and more easily repaired than RNA. As a result, DNA serves as a more stable carrier of genetic information that is essential to survival and reproduction. Two important functions evolved with the advent of the *DNA world*: transcription and replication. Transcription involves copying a gene of a DNA sequence to make an mRNA molecule for protein synthesis. Replication, on the other hand, is the process by which a double-stranded DNA molecule is copied to produce two identical DNA molecules. DNA replication is central to the reproduction of the first cells and the origin of life. The two biological entities, the cellular world and the virus world strongly interacted over the entire history of life, enriching each other's genetic makeup, but each retained its autonomy and lived in parallel lifestyles.

The invention of cell division marks the transition from the prebiotic world to the biotic world. The first cells underwent a vegetative cell division known as binary fission, where the genetic material was segregated equally into two identical daughter cells. Every time a cell divided, DNA polymerases helped duplicate the cell's gene, so that a copy of the original gene could be passed to each daughter cell. All cell divisions were preceded by a single round of DNA replication. Once life began, it was able to diversify and proliferate until it filled every niche on the planet.

There is no compelling reason for believing that life came to Earth from outer space. It thus seems reasonable, until proven otherwise, that life started on our planet. The cosmic and planetary conditions during the tail end of the heavy bombardment period formed the cradle of life. The 'hot-cradle' theory was first proposed by Woese [122] from molecular evidence. The Archean crustal evolution and the paleoecology of Earth's earliest fossil records favor impact-generated hydrothermal craters as the most likely cradle for life (4, 18-24). Such an environment

could have been a rich source of energy and of elements such as sulfur, iron, and phosphorous that have played an important role in biogenesis. Life arose through five hierarchical stages of increasing molecular complexity in the hydrothermal vent environment about 4 billion years ago [4]. Prebiotic synthesis began amidst a chaotic chemical mixture of cosmic ingredients as that present in the Murchison meteorite. Molecular selection at each hierarchical level offered a pathway by which smaller components could organize themselves into more complex molecules such as cell membranes, RNAs, proteins, and DNAs that became highly organized leading to the first cell. To paraphrase François Jacob [134], the origin of life does not produce innovation from the scratch, but owing to cumulative effects of hierarchical history on prebiotic synthesis. Nature functions by integration. It works on what already exists like a tinkerer who, during millions of years slowly fine-tuned the products step by step by natural selection that culminates in the emergence of the highly organized first cell. But once life had started, further evolution had to proceed mainly through mutation or the slight tinkering of already existing DNAs. But these were merely variations on previous cellular structures.

Most likely, the first life established was the hyperthermophilic microbial organisms around four billion years ago and began to evolve rapidly by Darwinian evolution. LUCA, the last universal common ancestor of life on Earth, was probably a single-cell hyperthermophilic bacterium-like organism, as suggested by recent molecular phylogenetics analyses [20]. LUCA's fingerprints are still visible in modern hyperthermophilic organisms that live in hydrothermal vents. LUCA is not thought to be the first cell, but the last before the divergence of bacteria and archaea [125].

Archean volcano-sedimentary rocks (4.0-2.5 Ga) host all the known types of biosignatures, including morphological fossils of cells, colonies, biofilms, or other biological

constructions, such as microbial mats and stromatolites. Various Archean microfossils from Canada, Greenland, Australia, South Africa, and India provide valuable clues to the distribution and nature of ancient microbial life [4]. These Archean biosignatures demonstrate that microbial life was abundant and possibly more diverse than currently believed.

Most likely, the hyperthermophilic bacteria gave rise to the evolution of anoxygenic bacteria as early as 3.4 Ga, and then oxygenic photosynthetic bacteria around 3.0 Ga (113). The emergence of oxygenic photosynthetic bacteria or cyanobacteria allowed life to escape the hydrothermal crater surface into the surface of the adjacent shallow ocean basin, which depended entirely on the solar power in the mesophilic environment. It was a major transition in the microbial life where oxygen began being toxic to becoming a vital part of metabolism. Cyanobacteria contributed to the geological processes by producing oxygen in our planet for the first time, and left their signature in shallow marine sediments by vast amounts of carbonate sediments and stromatolitic structure. Once cyanobacteria invaded the sea, they had global distribution. The invention of oxygenic photosynthesis by cyanobacteria forever transformed Earth. The biogeochemical shift set into motion the evolution of subsequent microbial metabolism and lifestyles. Recent fossil studies have provided evidence for the diversification of microbes in the Paleo- and Mesoarchean. Once life proliferated across the planet, it assumed remarkable forms, and wrought the extraordinary changes that have now inextricably linked to biosphere and geosphere.

Over the past decades, the boundary conditions under which microbial life can survive have been pushed in every possible direction, encompassing broader ranges of temperature, pH, pressure, salinity, energy, and nutrient limitation. Extremophiles are unicellular microbes (Bacteria, Archaea, and Eukaryotes) that can thrive in some of the most extreme environments

known today from Antarctic frozen lakes to continental hot springs and crater vents, submarine hydrothermal vents, dry desert soils, salt solutions, toxic wastes, organic solvents, heavy metals, radioactive waste, even 7 km-deep rocks inside earth's crust, as well as acid and alkaline niches that were previously considered inhospitable.

The study of these extremophiles has profound implications for the search of life on other planetary bodies. There is a strong possibility that life also emerged beyond Earth wherever the necessary physical and chemical conditions were met. 'Extra-terrestrial life' refers to a form of life resembling Earth life in all basic properties, including the DNA-RNA-protein triad [63]. Early life on Earth may inform us about what we might find on Mars, how it might be preserved, where to search for it, and how to confirm its biogenic origin. Using this information, we can search for potential life on other planetary bodies such as dwarf planet Ceres in the asteroid belt, Jupiter's moon Europa, and Saturn's moons Titan and Enceladus. Recently, the Kepler space telescope has discovered about 4,000 Earth-like exoplanets in our galaxy, each orbiting a Sun-like star in the 'Goldilocks zone' —that is, at just the right distance for liquid water. Some of these exoplanets may be habitable and may harbor life. NASA has detected water vapor and optimum temperatures in the atmosphere of the exoplanet, known as K2-18b that could potentially support life. The question is no longer, is there any life beyond Earth? The question is how do we detect life by remote sensing in our Solar System and beyond.

Funding: The research received no external funding.

Acknowledgments: I thank David Damer for inviting me to contribute this article in the special issue celebrating the five decades of pioneering research of David Deamer in the origin of life. I

thank Kippra Hopper for editing the manuscript for clarity and brevity and three anonymous reviewers and the editor for their helpful comments and constructive suggestions. I thank Volkan Sarigul for artwork and Texas Tech University for funding the research. The paper is dedicated to David Deamer who inspired me in the astrobiology research and provided critical information on the origin of cell membrane.

Conflicts of interest: The author declares no conflict of interest.

References

1. De Duve, C. The beginnings of life on earth. *Amer. Scient*., *83*, **1995**, 428-437.
2. Kasting, J. F. Atmospheric composition of Hadean-Early Archean Earth: The importance of CO. *Geol. Soc. Am. Spec. Paper* **2014,** *504*, 19-28.
3. Marchi, S.; Bottke, W.F.; Elkins-Tanton, L.T.; Bierhaus, M.; Wuennemann, K.; Morbidelli, A.; Kring, D.A. Widespread mixing and burial of Earth's Hadean crust by asteroid impacts. *Nature* **2014**, *511*, 578-582.
4. Chatterjee, S. The hydrothermal impact crater lakes: the crucibles of life's origin. In *Handbook of Astrobiology*; Kolb, V.M. Ed.; CRC Press; Taylor & Francis: Boca Raton, FL, USA, 2018; pp.265-295.
5. Mojzsis SJ, Harrison TM (2000) Vestiges of a beginning. Clues to the emergent biosphere recorded in the oldest known sedimentary rocks. *GSA Today*, **2000**, **10**, 1-5.
6. Chatterjee, S. The RNA/protein world and the endoprebiotic origin of life In *Earth, Life, and System*; Clarke, B. Ed; Fordham University Press, New York, **2015**, pp. 39-79.
7. Chatterjee, S. A symbiotic view of the origin of life at hydrothermal impact crater lakes. *Phys. Chem. Chem. Phys*. **2016**, *18*, 20033-20046.
8. Bernstein, M.P.; Sandford, S.A.; Allamonda, L.J. Life's far flung raw material. *Scient. Amer*. **1999**, *263*, 42-49.
9. Chyba, C.; Sagan, C. Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origin of life. *Nature* **1992**, *355*,125-132.
10. Deamer, D.W.; Dworkin, J.P.; Sandford, S.A.; Bernstein, M.P. Allamandola, L.J. The first cell membranes. *Astrobiol*. **2002**, *2*, 371-381.
11. Sandford, S.A.; Bernstein, M.P.; Dworkin, J.P. Assessment of the interstellar processes leading to deuterium enrichment in meteoritic organics. *Meteorit. Planet. Sci*., **2001**, *36*,1117-1133.
12. Westphal, A. J. et al. (with 67 authors). Evidence for interstellar origin of seven dust particles collected by the Stardust spacecraft. *Science* **2014**, *345*, 786-791.

13. Pizzarello, J.R.; Cronin, J.R. Non-racemic amino acids in the Murchison and Murray meteorites. *Geochem. Cosmochem. Acta* **2000**, *64*, 329-338.
14. Delsemme, A. H. Our Cosmic Origins. Cambridge University Press, Cambridge, UK, 1988. Hazen, R.M. *Genesis: The Scientific Quest for Life*. Joseph Henry Press: Washington, D.C., USA, 2005.
15. Deamer, D.W. *First Life: Discovering the Connections between Stars, Cells, and How Life Began*; University of California Press: Berkeley, CA, USA, 2011.
15. Boss, A.P.; Keiser, S.A. Who pulled the trigger: a supernova or an asymmetric giant branch star? *Astro. J. Lett.* **2010**, *717*, L1-L5.
17. Mojzsis, S.J.; Harrison, T.M.; Pidgeon, T.T. Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* **2001**, *409*, 178-181.
18. Cockell, C.S. The origin and emergence of life under impact bombardment. *Phil. Trans. R. Soc*. **2006**, *B361*, 1845-1856.
19. Osinski, G.R.; Tornabene, L.L.; Banerjee, N.R.; Cockell, C.S.; Flemming, R.; Izawa, M.R.M.; McCutcheon, J.; Parnell, J.; Preston, L.J.; Pickersgill, A.E.; Pontefract, A.; Sapers, H.M.; Southam, G. Impact-generated hydrothermal systems on Earth and Mars. *Icarus* **2013**, *224*, 347-363.
20. Weiss, M.C.; Sousa, F.L.; Mrnjavic, N.; Neukirchen, S.; Roettger, M.; Nelson-Sathi, S.; Martin, W.F. 2016. The physiology and habitat of the last universal common ancestor. *Nature Microbiol.*, **2016**, doi: 10.1038/NMICROBIOL,2016.116.
21. Stetter, K.O. Hyperthermophiles in the history of life. *Phil. Trans. R. Soc.*, **2006**, *361*, 1837-1843.
22. Baross, J.A.; Hoffman, S.E. Submarine hydrothermal vents and associated gradient environment as sites for the origin and evolution of life. *Origins of Life*, **1985**, *15*, 327-345.
23. Martin, W.F.; Sousa, F. L.; Lane, N. Energy at life's origin. *Science* **2014**, *344*, 1092-1093.
24. Kring, D.A. Impact events and their effect on the origin, evolution, and distribution of life. *GSA Today* **2000**, *10(8)*, 1-7.
25. Mulkidjanian, A.Y.; Bychkov, A.B.; Diprova, D.V.; Galperin, M.Y.; Koonin, E.B. Origin of first cells at terrestrial, anoxic geothermal fields. *Proc. Nat. Acad. Sci. USA*, **2012**,*109*, E821-E830.
26. Monnard, P.A.; Appel, C.L.; Kanavarioti, A., Deamer, D.W. Influence of ionic inorganic solutes on self-assembly and polymerization processes related to early form of life: Implications for a prebiotic aqueous medium. *Astrobiology*, **2002**, *2*, 139-152.
27. Mishteyn, D.; Damer, B.; Havid, J; Deamer, D. Amphiphilic compounds assemble into membranous vesicles in hydrothermal hot spring water but not in seawater. *Life*, **2018**, *8*, 11; 10.3390/life8020011.
28. Pearce, B.K.D; Pudriz, R.E.; Semenov, D.A.; Henning, T.K. Origin of the RNA world; The fate of nucleobases in warm little ponds. *Proc. Nat. Acad. Sci. USA*, **2017**,*114*, 11327-11332.
29. Tang, M.; Chen, K.; Rudnick, R.L. Archean upper crust transition from mafic to felsic marks the onset of plate tectonics. *Science*, **2016**, *351*, 373-375.
30. Cockell, C.; Lee, P. The biology of impact craters: a review. *Biol. Rev*. **2002**, *77*, 270-310.
31. Hagert, J.J.; Newsom, H. E. Hydrothermal alteration of the Lonar impact structure, India: Implications for impact cratering on Mars. *Meteor. Planet. Sci.*, **2003**, *38*, 365-381.

32. Parnell. J.; Lee, P.; Cockell, C.S.; Osinski, G.R. (2004) Microbial colonization in impact-generated hydrothermal sulphate deposits, Haughton impact structure, and implications for sulfate on Mars. *Int. J. Astrobiol.* **2004**, *3*, 247-25.
33. Grotzinger, J. P. et al. (2014) A habitable fluvio-lacustrine environment at Yellowknife Bay, Gale Crater, Mars. *Science*, **2014**, *343*, doi:10.1126/science 1242777.
34. Eigen, M.; Schuster, P. The hypercycle: a principle of natural self-organization. Part a: emergence of the hypercycle. *Naturwissens*, *64*, **1997**, 541-565.
35. Hazen, R.M. Genesis: The Scientific Quest for Life. Joseph Henry Press, Washington, DC, 2005.
36. Wachterhäuser, G. The cradle of chemistry of life: on the origin of natural products in a pyrite-pulled chemoautotrophic origin of life. *Pure Appl. Chem*. **1993**, *65*, 1343-1348.
37. Cech, T.R. RNA as an enzyme. *Scient. Amer*., **1986**, *255(5)*, 64-75.
38. Crick, F.H.C. The origin of genetic code. *J. Mol. Biol*., *38*, **1968**, 367-379.
39. Gilbert, W. The RNA world. *Nature*, *319*, **1986**, 618.
40. Orgel, L.E. The origin of life. *Scient. Amer*., **1994**, *271(4)*, 77-83.
41. Robertson, M.P.; Joyce, G.F. The origins of the RNA world. *Cold Spring Harb. Perspect Biol*., **2012**, *4(5)*, a003608.
42. 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.
43. Bowman, J.C.; Hud, N.V.; Williams, J.D. The ribosome challenge to the RNA world. *J. Mol. Evol.*, **2015**, *80*, 143-161.
44. Carter, C.W. Jr. What RNA world? Why a peptide/RNA partnership merits renewed experimental attention. *Life* **2015**, *5*, 294-320; doi:10.3390/life5010294.
45. Carter, C.W. Jr. An alternative to the RNA world. *Nat. Hist*. **2016**, *125(1)*, 28-33.
46. Carter, C. W. Jr.; Wills, P. R. Interdependence, reflexivity, impedance matching, and the evolution of genetic coding. *Mol. Biol. Evol*. **2017**,
47. Harish, A.; Caetano-Anolles, G. Ribosomal history reveals origin of modern protein synthesis. *PLoS One*, **2001**, doi:e32776.
48. Dyson. F. *Origins of Life*. Cambridge University Press: Cambridge, UK, 2004.
49. Leu, K.; Obermayer, B.; Rajamani, S.; Gerland, U.; Chen, I.A. The prebiotic evolutionary advantage of transferring genetic information RNA to DNA. *Nucleic Acid Res.* **2011**,1-13; doi: 10.1093/nar/gkr525.
50. Deamer, D.W. The first living systems: a bioenergetic perspective. *Microbiol. Mol. Biol. Rev*.**1997**, *61*, 230-261.
51. Szostak, J.W.; Bartel, D.P., Luisi, P.L. Synthesizing life. *Nature*, **2001**, *409*,387-390.
52. Deamer, D.W. The role of lipid membrane in life's origin. *Life*, **2017**, *7(1)*, 5; https://doi.org/10.3390/life3701005.
53. Mansy, S.S., Schrum, I.P.; Krishnamurthy, M., Tobe, S., Treco, D.A.; Szostak, J.W. 2008. Template-directed synthesis of a genetic polymer in a model protocell. *Nature*, **2008**, *454*, 122-126.
54. Mansy, S.S. Model protocells from single-chain lipids. *Int. J. Mol. Biol*., **2009**, *10*, 835-843.
55. Deamer, D.W. *Assembling Life: How Can Life Begin of Earth and Other habitable Planets?* Oxford University Press, New York, 2019.

56. Deamer, D.W.; Pashley, R.M. Amphiphilic components of the Murchison carbonaceous chondrite: Surface properties and membrane forming formation. *Orig. Life. Evol. Biosph.*, **1989**, *19*, 21-38.
57. Dworkin, J; Deamer, D.; Sandford, S.; Allamandola, I. Self-assembling amphiphilic molecules: Simulated in simulated interstellar/precometary ices. *Proc. Natl. Acad. Sci.*, **2001**, *98*, 825-819.
58. McCollom, T.M., Ritter, G.; Simonetta, B.R. Lipid synthesis under hydrothermal conditions by Fischer-Tropsch-type reactions. *Orig. Life Evol. Biosph*., **2003**, 29, *153*-166.
59. Monnard, P.A.; Deamer, D.W. Membrane self-assembly processes: steps toward the first cellular life. *Anat. Rec*., **2002**, *268*, 197-207.
60. Damer, B.; Deamer, D.W. Coupled phases and combinatorial selection in fluctuating hydrothermal pools: a scenario to guide experimental approaches to the origin of cellular life. *Life* **2015**, *5*, 872-887.
61. Zhu, T.F.; Szostak, J.W. Coupled growth and division of model protocell membranes. *J. Am. Chem. Soc*., **2009**, *131*, 5705-5713.
62. Chen, I.A.; Szostak, J.W. A kinetic study of the growth of fatty acid vesicles. *Biophys. J.* **2004**, *87*, 988-998.
63. De Duve, C. *Singularities: Landmarks on the Pathways of Life*: Cambridge University Press, New York, USA, 2005.
64. Budin, I.; Szostak, J.W. Physical effects underlying the transition from primitive to modern cell membranes. *Proc. Nat. Acad. Sci*., **2011**, *108*, 549-5254. http://dx.doi.org/10.1093/molbev/msx265.
65. Hargraves, W.W.; Mulvihill, S.J.; Deamer, D.W. Synthesis of phospholipids and membranes in prebiotic conditions. *Nature*, **1977**, *266*, 78-80.
66. Apel, C.L.; Deamer, D.W. The formation of glycerol monodecanoate by a dehydration condensation reaction: The chemical complexity of amphiphiles on the early Earth. *Orig. Life. Evol. Biospheres*, **2005**, *35*, 323-332.
67. Simoneit, B.R.T.; Rushdi, A.I.; Deamer, D.W. Abiotic formation of acylglycerols under simulated hydrothermal conditions and self-assembly properties of such lipid products. *Adv. Space Res*., **2007**, *40*, 1649-1656.
68. Singer, S.J.; Nicholson, G.L. The fluid mosaic model of the structure of the cell membrane. *Science*, **1972**, *175*,720-731.
69. La Scola, B.; Audic, S.; Robert, C.; Jungang, L.; Lamballerie. X. D.; Drancourt, M.; Birtles, R.; Claverie, J.M.; Raoult, D. A giant virus in Amoebae. *Science*, **2003**, *299*, 2033.
70. Forterre, P. Defining life: The virus viewpoint. *Orig. Life. Evol. Biosph*., **2010**, *40*, 151-160.
71. Koonin, E.V.; Senkevich, T.G.; Dojla, V.V. The ancient virus world and evolution of cells. *Biol. Dir.* **2006**, *1:29*, doi: 10.1186/1745-6150-1-29.
72. Suttle, C.A. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.*, **2007**, *5*, 801-812.
73. Villarreall, L.P. *Viruses and Evolution of Life*; American Society of Microbiology Press, Washington, DC, 2005.
74. Wagner, E.K; Hewlett, M.J. *Basic Virology*; Blackwell Publishing; Malden, MA, USA, 2004.

75. Koonin, E.V. Viruses and mobiles elements as drivers of evolutionary transitions. *Phil. Trans. R. Soc. B*, **2016**, *371*, http://dx.doi.org/10.1098/rstb.2015.0442.
76. Villarreall, L.P. Are viruses alive? *Scient. Amer.*, **2004**, *291(1)*, 100-105.
77. D'Herelle, F. *The Bacteriophage: Its Role in Immunity*. Williams and Wilkins, Baltimore, 1922.
78. Haldane, J.B.S. The origin of life. *Ration. Ann.* **1929**, 148-169.
79. Forterre, P. The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. *Biochemie.*, **2005**, *87*, 793-803.
80. Margulis, L. Symbiosis in Cell Division. Freeman, San Francisco, CA, 2005.
81. Takemura, M. Poxyviruses and the origin of the eukaryotic nucleus. *Mol. Evol. Biol.*, **2001**, *52*, 419-425.
82. Chatterjee, S. The river of life: a genetic perspective on macroevolution. Forum on Public Policy *Journal of Oxford Round Table*, **2009**, *5(5)*, 1-43.
83. Doolittle, W.F. (2000) Uprooting the tree of life. *Scient. Amer.*, **2000**, *282(2)*, 90-95.
84. Daubin, V; Ochman, H. Start-up entities in the origin of new genes. *Curr. Opin. Genet. Dev.* **2004**, *14*, 616-619.
85. Feschotte, C. Virology: bornavirus enters the genome. *Nature*, **2010**, *463*, 39-40.
86. Chuong E.B. Retroviruses facilitate the rapid evolution of the mammalian placenta. *Bioessays*, **2013**, *35(10)*, doi:10.1002/bies.201300059.
87. Holmes, E.C. What does virus evolution tell us about virus origin? *J. Virol.*, **2011**, *85*, 5247-525.
88. Forterre, P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.*, **2006**, 117, 5-16.
89. Koonin, E.V. On the origin of cells and viruses: Primordial virus world scenario. *Ann. N.Y. Acad. Sci.*, **2009**, *1178*, 47-64.
90. Koonin, E.V.; Dojla, V.V. A virocentric perspective on the evolution of life, *Curr. Opion. Virol.*, **2013**, 3(5), 546-557.
91. Koonin, E.V.; Dojla, V.V. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiol. Mol. Biol. Rev.*, **2014**, *78*, 278-303.
92. Krupovic, M.; Koonin, E.V. Multiple origins of viral capsid proteins from cellular ancestors. *Proc. Nat. Acad. Sci.*, **2011**, doi.10.1073pnas.1621061114.
93. Filee, J.; Forterre, P. Viral proteins functioning in organelles: a cryptic origin? *Trends. Microbiol.* **2005**, *13*, 510-513.
94. Durzynska, J.; Gozdzicka-Jozefiak, A. 2015. Viruses and cells intertwined since the dawn of evolution. *Virol. J.* **2015**, 12:169; doi: 10.1186/s12985-0400-7.
95. Sabbath, N.; Wagner, A.; Karlin, D. Evolution of vital proteins originated de novo by overprinting. *Mol. Biol. Evol.*, **2012**, *29*, 3767-3780.
96. Ahlquist, P; Noueiry, A.O.; Lee, W.M.; Kushner, D.B.; Dye, B.T. Host factors in positive-strand RNA virus genome replication. *J. Virol.*, **2003**, *77*, 8181-8186.
97. Brierly, I. Ribosomal frameshifting on viral RNAs. *J. Gen. Virol.* **1995**, *76*, 1885-1892.
98. Krupovic, M; Prangishvilli, D.; Hendrix, R.W.; Bamford, D.H.; Genomic of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol. Mol. Biol. Rev.*, **2011**, *75*, 610-635.
99. Diemer, G.S.; Stedman, K.M. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biol. Direct*, **2012**, *7*,13, doi: 10.1186/1745-6150-7-13.

100. Aiewsakun, P.; Katzourakis, A. Marine origin of retroviruses in the Early Paleozoic Era. *Nature Com.* **2017**, 8:13954, doi:10.1038/ncomms133954.
101. Forterre, P. Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc. Nat. Acad. Sci*., **2006**, *103*, 3669-3674.
102. Powner, M.W.; Zheng, S.L.; Szostak, J.W. Multicomponent assembly of proposed DNA precursors in water. *J. Amer. Chem. Soc*., **2012**,*134*,13889-13895.
103. Forterre, P.; Prangishvilli, D. The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as major source of evolutionary novelties. *Ann. N.Y. Acad. Sci*., **2009**, *1178*, 65-77.
104. Xu, J.; Green, N.J.; Gibard, C., Krishnamurthy, R.; Sutherland, J.D. Prebiotic phosphorylation of 2-thiouridine provides either nucleotides or DNA building blocks via photoreduction. *Nature Chem*. **2019**, *11*, 457-462.
105. Takeuchi, N.; Hogeweg, P.; Koonin, E.V. 2011. On the origin of DNA genomes: Evolution and division of labor between template and catalyst in model replicator systems. *PLoS Comp. Biol.*, **2011**, *7*, 3 e1002024.
106. Freeman, S. *Biological Science*, 2nd edition, Pearson Prentice Hall: Upper Saddle River, New Jersey, USA, 2005.
107. Watson, J.D.; Crick, F.H.C. A structure for deoxyribose nucleic acid. *Nature*, **1953**, *171*, 737-738.
108. Schrödinger, E. *What is Life? The Physical Aspects of Living Cell*. Cambridge University Press: Cambridge, UK,1944.
109. Catling, D.C. Astrobiology: A Very Short History of Introduction. Oxford University, Press, Oxford, 2013.
110. Von Neumann, J. *Theory of Self-Reproducing Automata*. University of Illinois Press: Chicago, IL, USA, 1966. Edited and completed by A.W. Burks.
111. Walker, S.I.; Davies, P.C.W. The algorithmic origins of life. *J. R. Soc. Interface*, **2012**, doi: 10.1098/rsif.2012.0869.
112. Maturana, H.R.; Valera, F.J. Autopoiesis and Cognition. The Realization of Living. *Boston Stud. Phil. Sci., 42*, D. Reidel Publishing: Boston, 1981.
113. Schopf, J.W. *Cradle of life: The Discovery of Earth's Earliest Fossils*. Princeton University Press: Princeton, New Jersey, 1999.
114. Goodsell, D.S. *The Machinery of Life.* Springer, New York, 2010.
115. Chen, I.A. Cell division: breaking up is easy to do. *Curr. Biol*., **2009**, *19*, R327-R328.
116. Zwicker, D; Seyboldt, R.; Weber, C.A.; Hyman, A.A.; Jülicher, F. Growth and division of active droplets provides a model for protocells. *Nature Phys*., **2017**, *13*, 408-413.
117. Kurihara, K.; Okura, Y.; Matsuo, M.; Toyota, T.; Suzuki, K.; Sugawara, T. A recursive vesicle-based model protocell.
118. Saha, R.; Chen, I.A. Origin of life: protocells red in tooth and claw. *Curr. Biol*., **2015**, *25*, R1166-R1185.
119. Leaver, M.; Domingez-Cuevas, P.; Coxhead, J.M.; Daniel, R.A.; Erington, J. Life without a wall or division machine in *Bacillus subtillis*. *Nature*, **2009**, *457*, 849-853.
120. Panno, J. *The Cell: The Nature's First Cell Life Form.* Facts on File: New York. 2010.
121. Darwin, C. *On the Origin of Species*. John Murray, London, 1859.
122. Woese, C.R. On the evolution of cells. *Proc. Nat. Acad. Sci*., **2002**, *99*, 8742-8747.

123. Williams, T.A.; Foster, P.G.; Cox, C.J.; Embley, T.M. An archaeal origin of eukaryotes supports only two primary domains. *Nature*, **2013**, *504*, 231-236.
124. Woese, C.R. The universal ancestor. *Proc. Nat. Acad. Sci*., **1994**, *99*, 6864-6859.
125. Cornish-Bowden, A.; Cardenas, M.L. Life before LUCA. *Jour. Theor. Biol*., **2017**, *434*, 68-74.
126. Theobald, D. L. A formal test of the theory of universal common ancestry. *Nature*, **2010**, *465*, 219-223.
127. Gogarten, J.P.; Deamer, D. Is LUCA a thermophilic progenote? *Nature Microbiol*. **2016**, doi: 10.1038/NMICROBIOL2016229.
128. Akanuma, S.; Nakajima, Y.; Yokobori. S.; Kimura, M.; Nemoto, N.; Mase, T.; Miyazono, K. Tanokura M, Yamagishi A (2013) Experimental evidence for the thermophilicity of ancestral Life. *Proc. Nat. Acad. Sci.*, **2013**, *110*, 11067-11072.
129. Gaucher, E.A.; Govindan, S.; Ganesh, O.K. Paleotemperature trend for Precambrian life inferred from resurrected proteins. *Nature*, 2008, *451*, 704-707.
with a primitive model cycle. *Nature Comm*., **2015**, doi: 10.1038/ncomms9352.
130. Pace, N.R. A molecular view of microbial diversity and the biosphere. *Science*, **1997**, *276*, 734-740.
131. Reysenbach, A.L.; Cady, S.L. Microbiology of ancient and modern hydrothermal systems. *Trend Microbiol.*, 2001, 9, 79-86.
132. Nguyen, V.; Wilson, C.; Hoemberger, M.; Stiller, J.B.; Agavonof, R.V.; Kutter, S.; English, J.; Theobald, D.L., Kern, D. Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science*, **2017**, *355*, 289-294.
133. Nisbet, E. G.; Sleep, N. H. The habitat and nature of early life. *Nature*, **2001**, *409*, 1083-1091.
134. Jacob, F. Evolution and tinkering. *Science*, **1977**, *196*, 1161-1166.

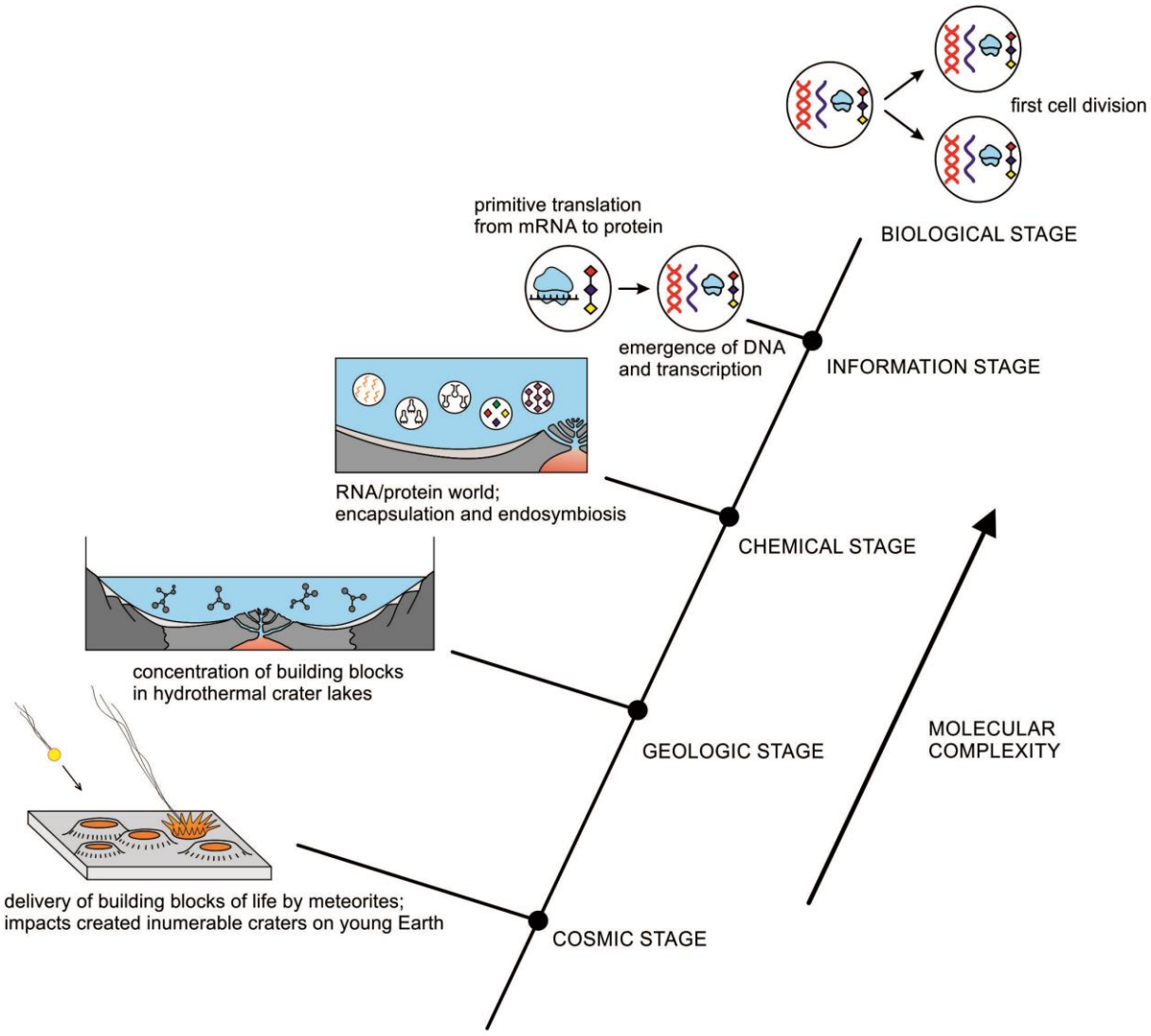


Figure 1. The hierarchical origin of life, viewed as five ascending stages of increasing complexity, showing the biomolecules in the prebiotic world that led to the development of the first cells. These are the cosmic, geological, chemical, informational, and finally, biological stages. Each higher level provided novel emergent properties not found at any lower level. These sequential steps took place in the dark hot environments of hydrothermal crater lake basins, which may well represent the Earth's oldest ecosystems.

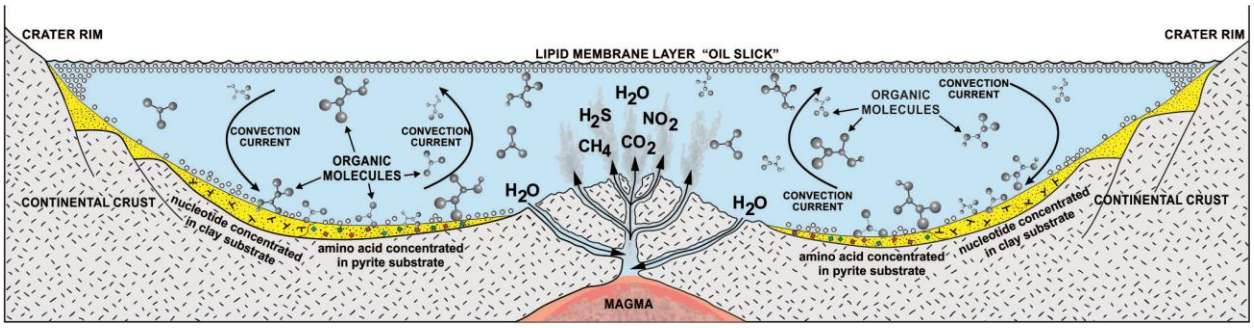


Figure 2. Cradles of life. During the Early Archean period (~4 Ga), freshwater crater basins with hydrothermal vent systems at their central peaks offered a protective sanctuary for the origin of life. The boiling water was rich with organic molecules brought by meteorites. On the water's surface, primitive lipid membranes and hydrocarbons floated like oil slicks. The minerals on the floors of the basin acted as catalytic surfaces for the concentration and polymerization of monomers. The bubbling biotic soup was thoroughly mixed by convection currents. These same currents also circulated some of the lipid membranes down to the basin floor where they attached to the porous mineral layers, encapsulating biopolymers such as RNA and amino acids. Heat, gases, and chemical energy, including ATP molecules released from the hydrothermal vents, brewed and condensed the prebiotic soup, which began to collect on the mineral substrate at the bottom of the basin (modified from Chatterjee 2016).

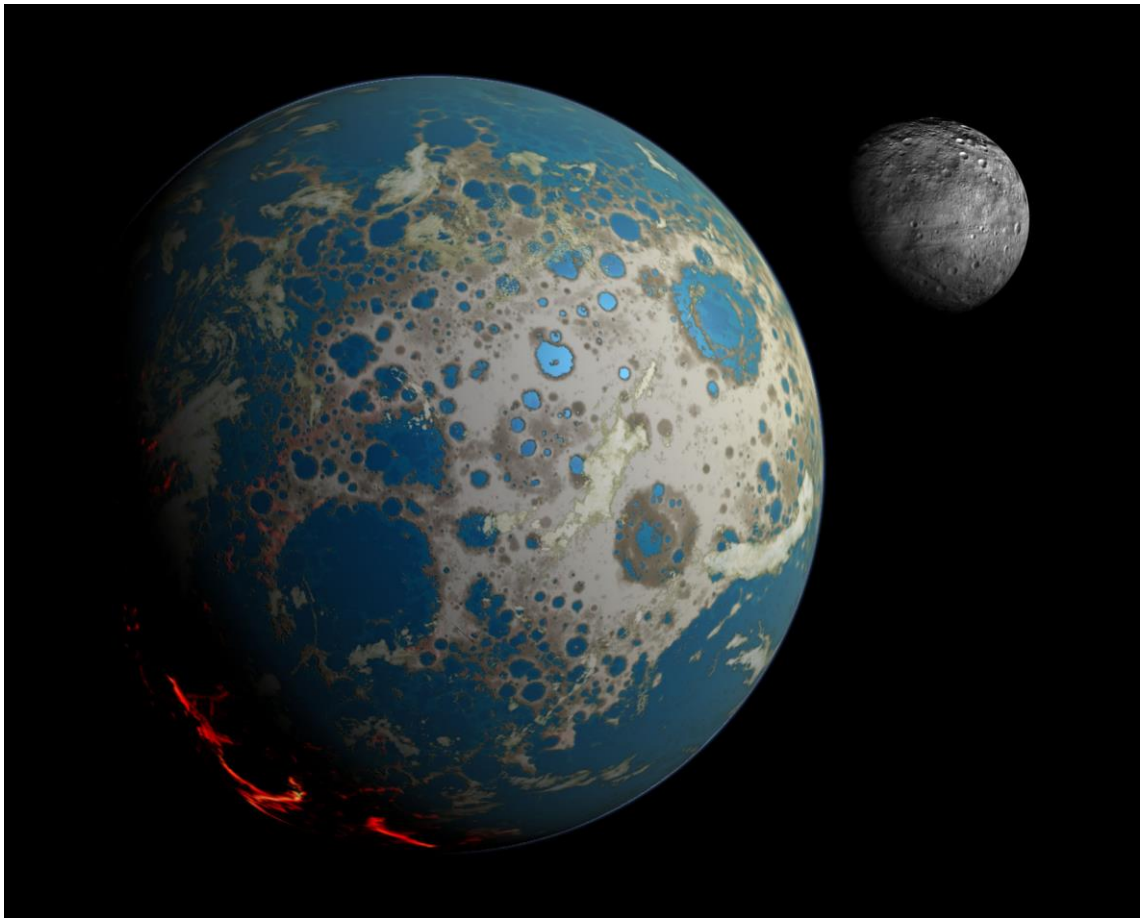


Figure 3. An artistic concept of the early Earth–Moon system (~4 Ga). The surface of the Hadean/Archean Earth was continually reprocessed by impacts, which mixed and buried the impact-generated melt. The Moon is shown as a dry, heavily cratered body, far less geologically active than the Earth because it lacks plate tectonics and a dynamic atmosphere. The crater records of the Moon have been used to calibrate the Late Heavy Bombardment period on Earth (Marchi et al., 2014; courtesy Simone Marchi).

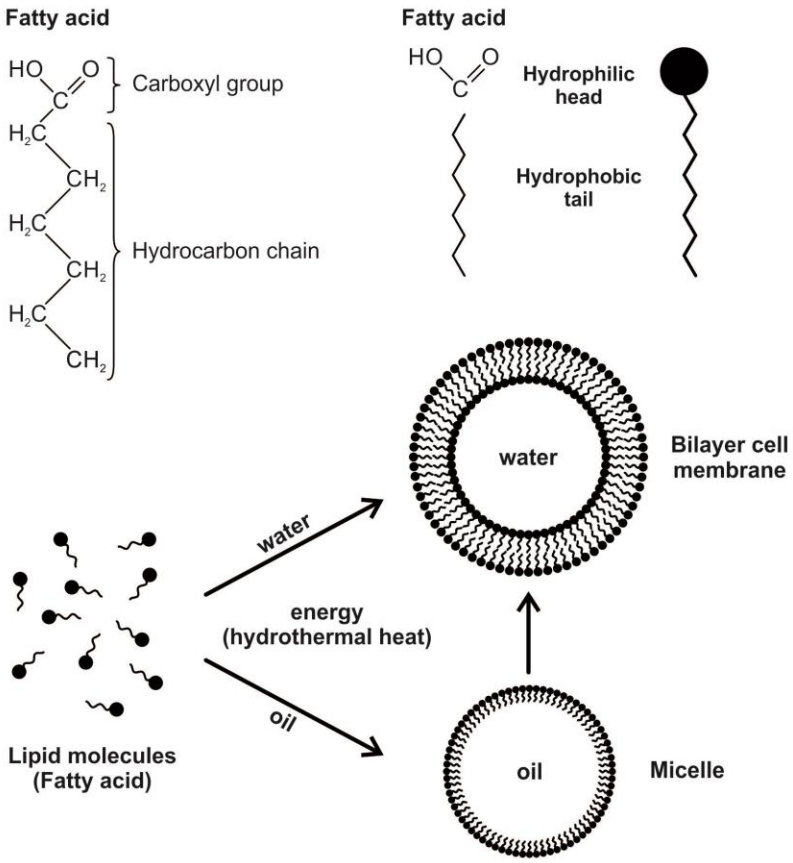


Figure 4. A simple fatty acid may have been a major component of the early prebiotic cell membrane. It has hydrophilic head and hydrophobic tail and can form vesicles of monolayer or bilayer. The monolayer vesicle (micelle) can only trap oils, not water, and cannot be precursor to the cell. On the other hand, the bilayer vesicle can trap water and water-soluble molecules and can give rise to the cell membrane.

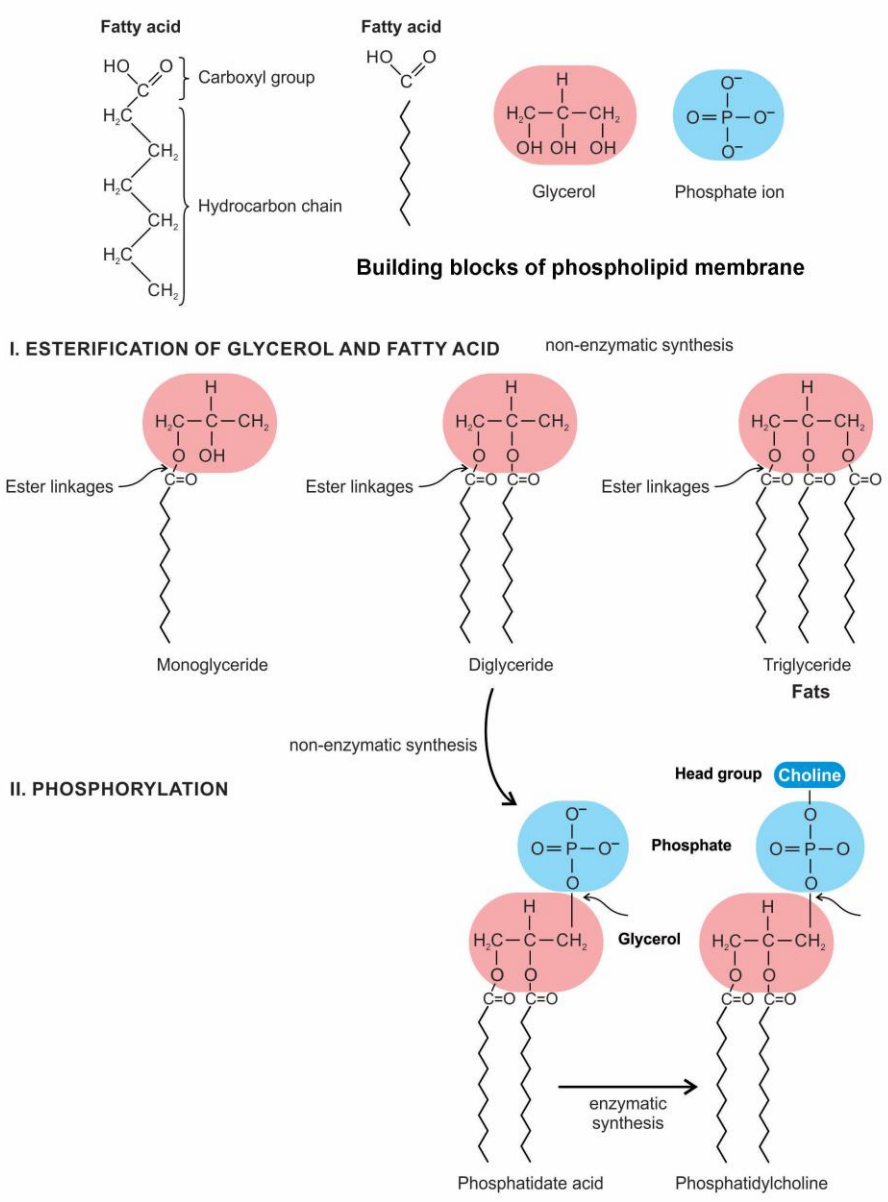


Figure 5. Origin of phospholipid membrane from simple fatty acids by an intermediate, the phosphatidate acid by a series of non-enzymatic synthesis. Fatty acids, glycerol, and phosphate ion are the building blocks of phospholipid. In the first stage, several glycerides such as monoglyceride, diglyceride, and triglyceride (fat) are formed by esterification of glycerol and fatty acid, with the loss of a water molecule; the covalent bond, an ester linkage results from this reaction. The next stage of synthesis of phosphatidate acid is by phosphorylation of a diglyceride molecule, when a phosphate ion is joined. Phosphatidate acid, in turn, would give rise to phospholipid by attaching to an alcohol molecule such as choline, ethanolamine, serine or inositol. Of these various combinations, phosphatidylcholine (shown in the figure) is the most common phospholipid in cell membrane.

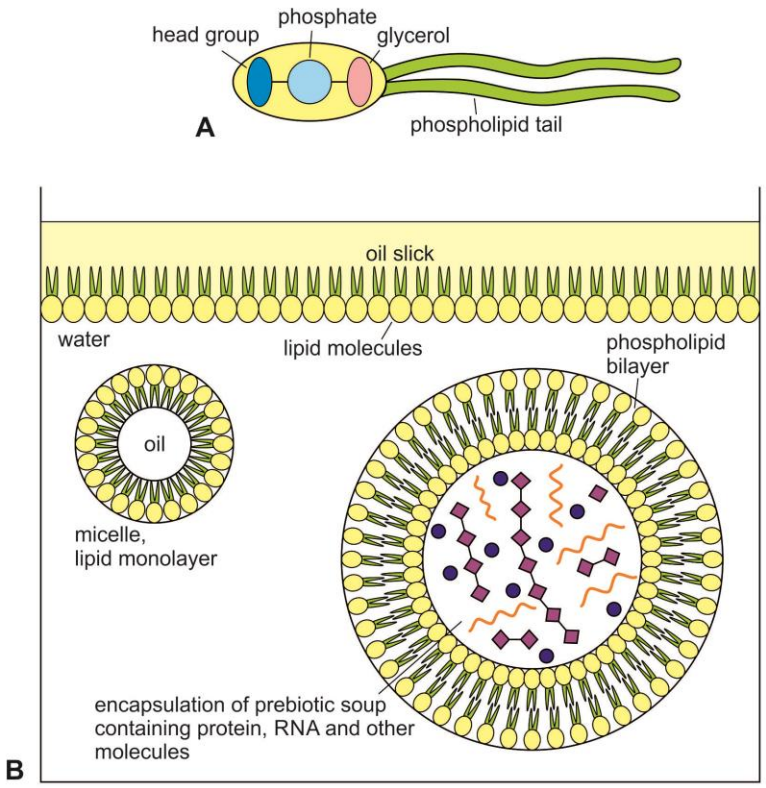


Figure 6. Self-assembly of phospholipid membrane in a hydrothermal crater lake in the peptide/RNA world. A, a generalized phospholipid molecule has a hydrophilic ('water-loving') head and two hydrophobic ('water-hating') tails that do not mix with water and will avoid being surrounded by it. B, in an oil slick on the surface of a crater lake, the hydrophobic tails mix with the oil. whereas the heads stay close to the water. During turbulence, phospholipids form two kinds of membranes: a monolayer, which can only capture a drop of oil (left), or a bilayer, which can capture a group of water molecules (right). The bilayer allows the hydrophobic tails to associate with one another, whereas the heads associate with water molecules, on both the inside and outside surfaces of the membrane. A bilayer vesicle is stabilized when it encapsulates protein molecules that interact with the bilayer surface. C, the formation of a semicell on a pyrite mineral surface, which allowed the encapsulation of protein molecules, leading to an endosymbiosis. The phospholipid membrane protocell provided a safe haven for the encapsulated protein molecules. In the protein/RNA world, as more proteins were synthesized by translation, they contributed molecules to the phospholipid membrane. These strengthened the cell wall, and made the passage of sodium and potassium ions and other nutrients through the membrane more efficient. This created a new plasma membrane, which operated like a protein–lipid 'sandwich.' Along with plasma membrane, the primitive cytoplasm also evolved as more and proteins were synthesized that converted the aqueous solution inside protocell to a gel-like substance. Left top, a cross-section of the plasma membrane shows its 'sandwich' reinforcement by protein molecules, which act as gatekeepers, pumping molecules in and out of the cell membrane.

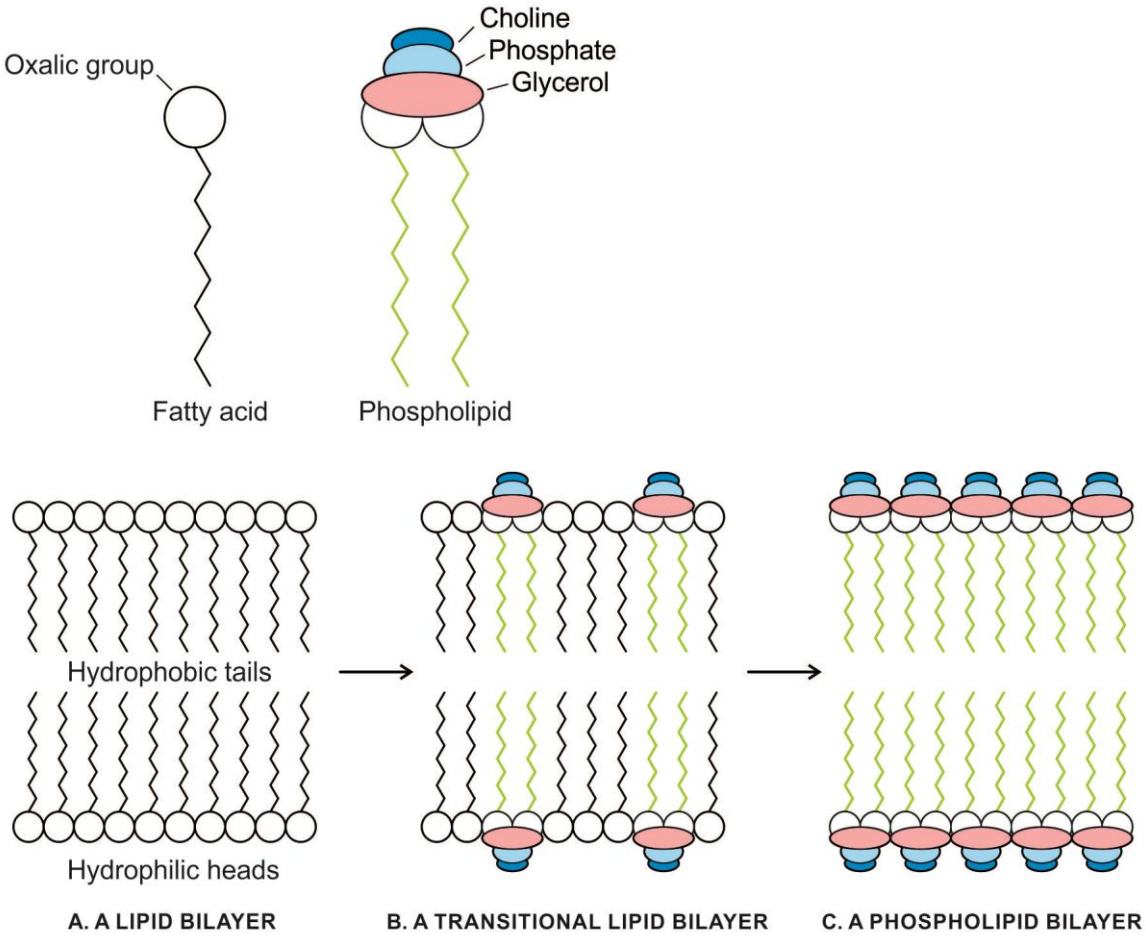


Figure 7. Gradual transition from single-chain, highly permeable fatty acid membrane to selective permeable phospholipid membrane with increase of phospholipid content via a transitional stage. Increasing phospholipid content inhibits the permeability of fatty acid membranes through changes in bilayer fluidity.

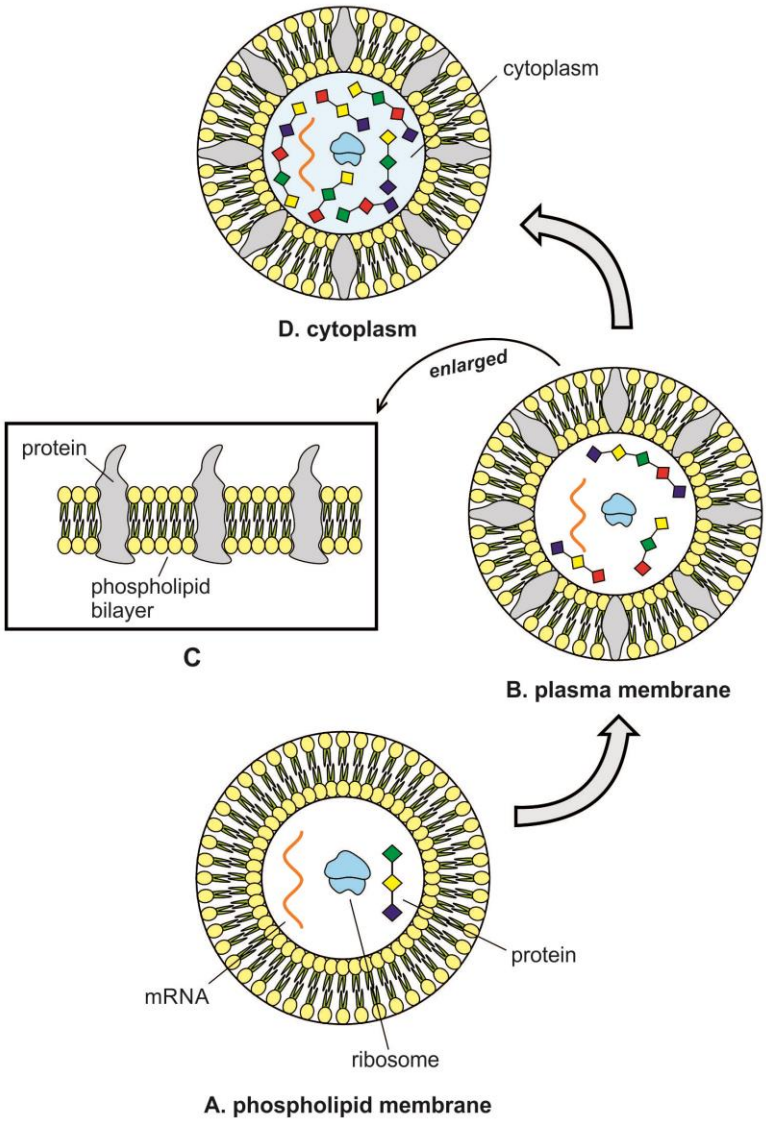


Figure 8. Transition of phospholipid membrane to plasma membrane. A, phospholipid membrane; B, plasma membrane; C, phospholipid membrane evolved into plasma membrane by inserting protein molecules into bilayers that made the cell membrane selective permeable so that certain ions like potassium and sodium can cross the bilayer barrier; D, cytoplasm inside protocell formed when newly synthesized protein molecules by translation machine began to concentrate inside protocell, turning internal water into a gel-like substance.

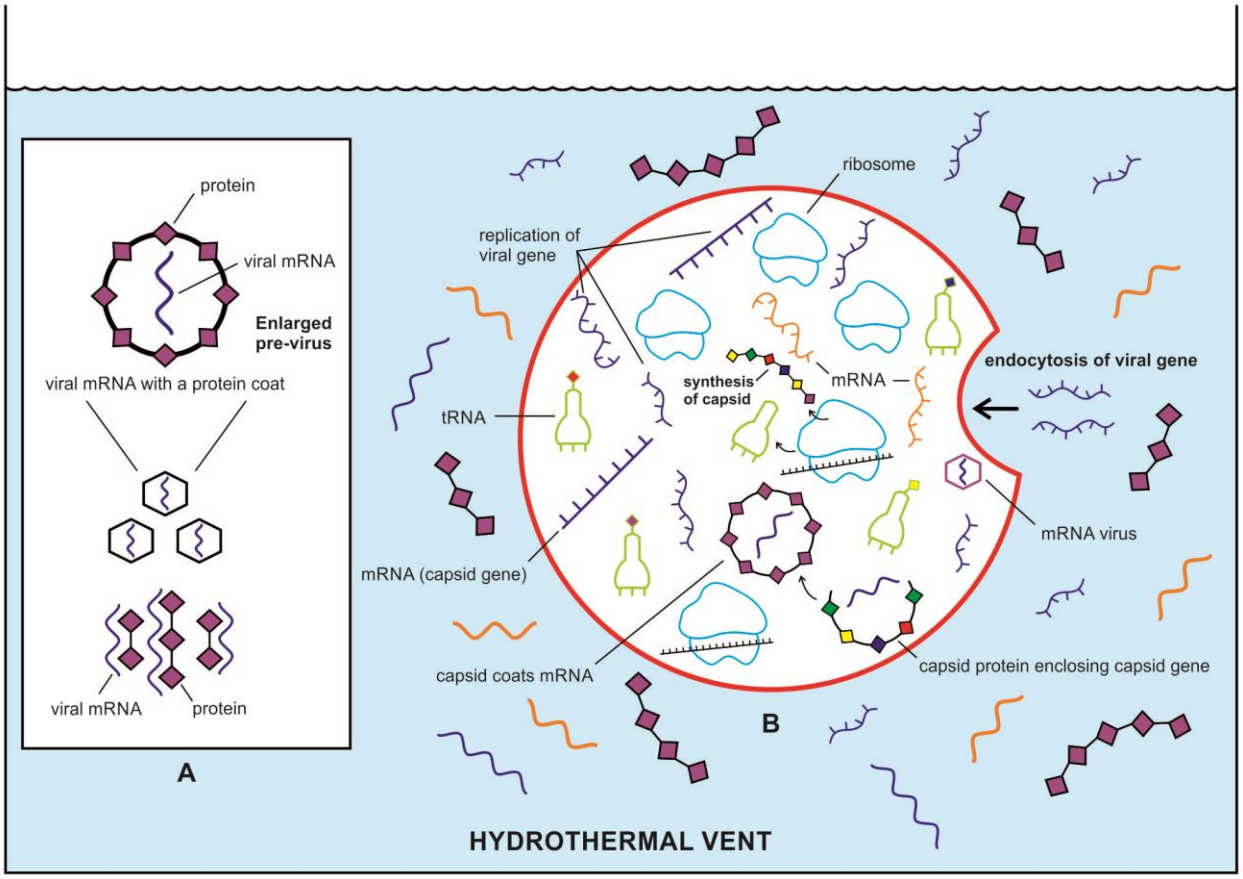


Figure 9. Beginnings of the ancient viral world in the prebiotic soup in the hydrothermal vent environment, which was an ideal Nature's genomic laboratory. A, in our model, viral selfish genes modified from mRNA molecules appeared de novo by overprinting in the prebiotic soup; these capsid genes were encased by a protein shell for stability and durability. However, these proteins were random, easily available in the prebiotic soup, and were not encoded by viral gene. This initial stage of viral structure that afforded protection of fragile mRNA is called 'pre-virus'. B, In the next stage of evolution, some viral genes entered the protocells by endocytosis, utilizing their ribosomes for synthesis of capsid proteins. Once the capsid protein began to coat viral gene, the first mRNA virus appeared.

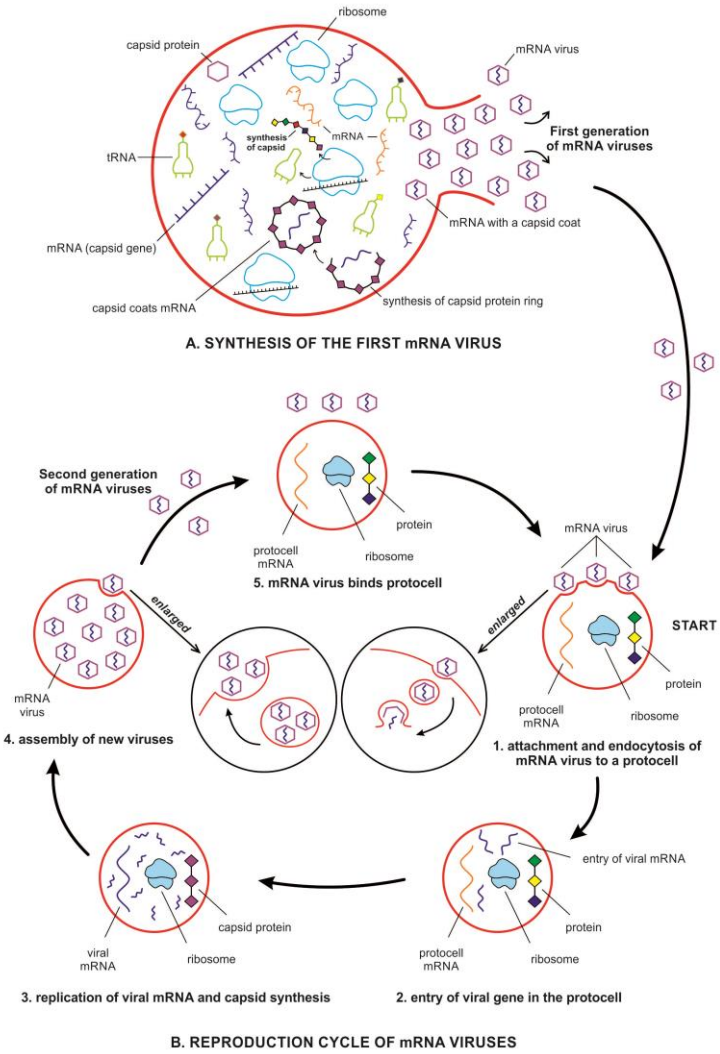


Figure 10. Origin of mRNA viruses and reproduction cycle**.** A, synthesis of first generation of mRNA viruses inside protocells hijacking hosts' translation machinery. mRNA viral genes produced capsid proteins that began to wrap viral genes. As more and more mRNA viruses were synthesized inside protocells, they exerted osmotic pressure on the phospholipid membrane causing a burst of protocells, releasing the first batch of viruses for infection of protocells. B, some likely stages of the reproduction cycle of mRNA virus includes: attachment of a protocell and its entry into protocell via endocytosis (cycle 1), uncoating of capsid shell (cycle 2), mRNA copying and protein synthesis (cycle 2 and 3), self-assembly of viral progeny and release via exocytosis (cycle 4).

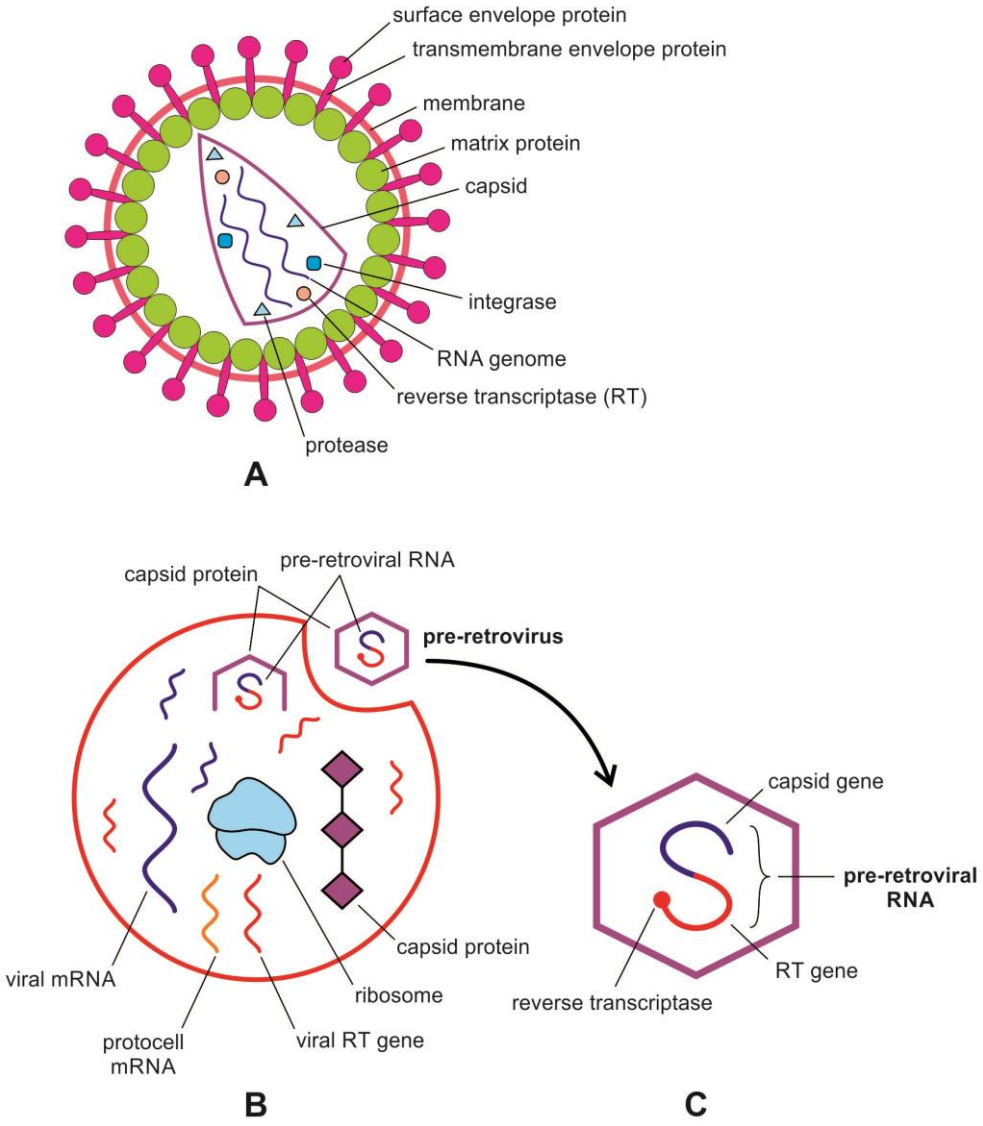


Figure 11. Retroviruses. A, structure of a modern retrovirus, an enveloped particle in which the capsid core contains two identical single-stranded mRNA molecules. Each mRNA is made up of three genes: integrase, reverse transcriptase, and protease. Once inside the host cell cytoplasm, the virus uses its own reverse transcriptase enzyme to produce DNA from its RNA genome. B, likely origin of pre-retrovirus inside infected protocell, where the viral RT gene was linked to the mRNA viral gene, and was encased by a capsid coat. Once two genes were fused into a single gene for close packing, pre-retrovirus released from the protocell via exocytosis. It was a non-enveloped particle.

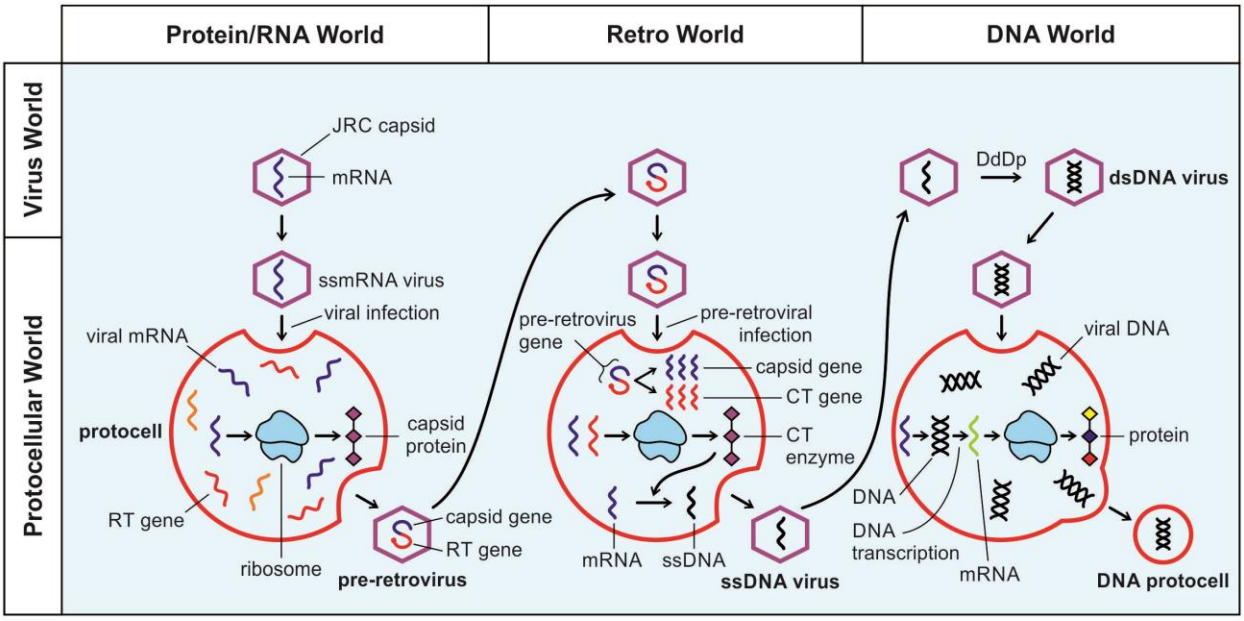
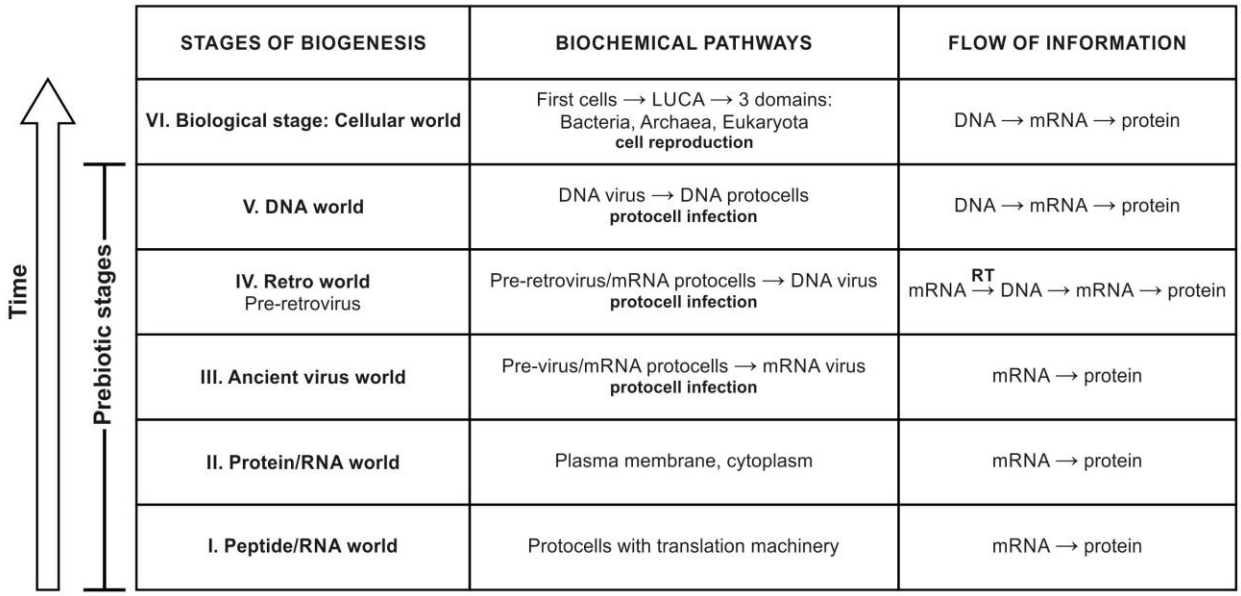


Figure 12. Postulated evolution of the biochemical pathways from the peptide/RNA world to the DNA world and the concurrent building of the information system to contemporary central dogma (DNA-makes-RNA-makes-proteins).



STAGES OF BIOGENESIS	BIOCHEMICAL PATHWAYS	FLOW OF INFORMATION
VI. Biological stage: Cellular world	First cells → LUCA → 3 domains: Bacteria, Archaea, Eukaryota **cell reproduction**	DNA → mRNA → protein
V. DNA world	DNA virus → DNA protocells **protocell infection**	DNA → mRNA → protein
IV. Retro world Pre-retrovirus	Pre-retrovirus/mRNA protocells → DNA virus **protocell infection**	mRNA $\xrightarrow{\text{RT}}$ DNA → mRNA → protein
III. Ancient virus world	Pre-virus/mRNA protocells → mRNA virus **protocell infection**	mRNA → protein
II. Protein/RNA world	Plasma membrane, cytoplasm	mRNA → protein
I. Peptide/RNA world	Protocells with translation machinery	mRNA → protein

Figure 13. Coevolution of viruses and protocells. Viruses became important vectors for donating critical enzymes and modifying genomes of the protocells during recurrent infection. Later, some of those viruses evolved DNA as a way to defend their genomes from attack, and DNA-based viruses became incorporated into hosts. In between are the fundamental catalytic processes that allowed to stepwise generate viral deoxyribonucleotides from ribonucleotides by RNA polymerase (RdRp), reverse transcriptase (RT), and DNA polymerase (DdRp) enzymes.

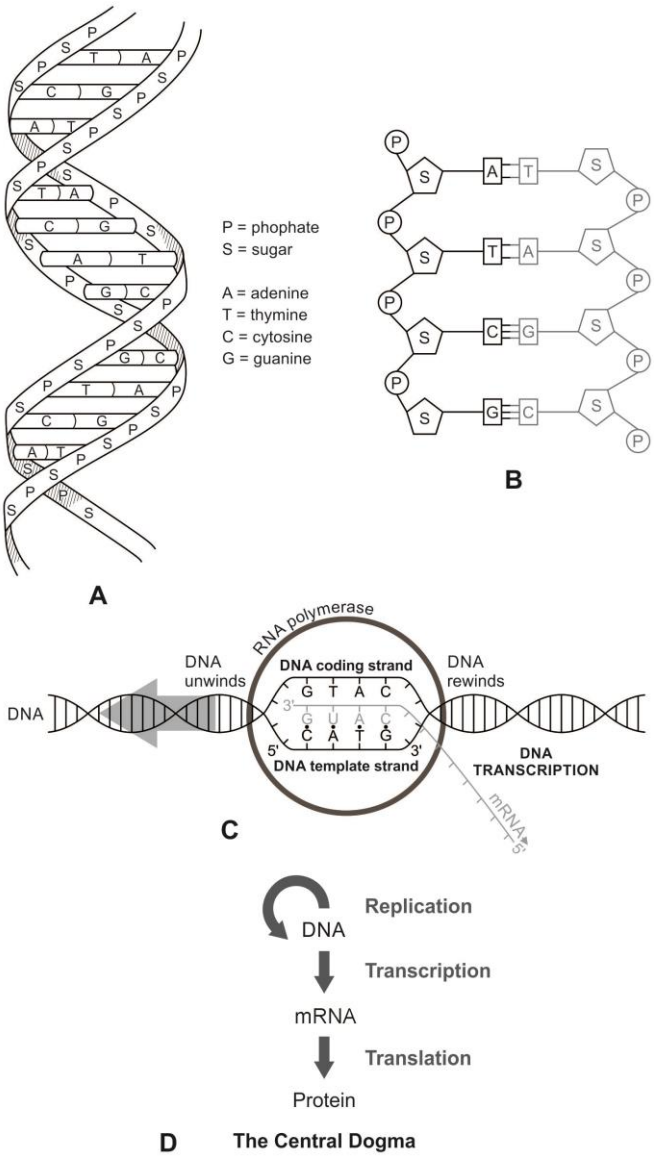


Figure 14. DNA structure and transcription. A, the twisted ladder model of the double helix of DNA showing two antiparallel strands, which are held together by hydrogen bonds between their bases. B, unwounded ladder structure of DNA showing the sides of the ladder made sugar-phosphate backbones of two strands; the four bases joined by hydrogen bonds forming the rungs of the ladder. C, DNA transcription; C, protein synthesis begins when a region of DNA is teased apart, and a molecule of mRNA is built along one template strand by an enzyme called 'RNA polymerase'. When the mRNA transcript is formed, it peels away from the DNA, allowing the already transcribed DNA to rewind into a double helix. The mRNA conveys the genetic information from the DNA to the ribosome, where it specifies a sequence of amino acids that will form a particular protein, in a process called translation. tRNA molecules bring the appropriate amino acids to assemble the protein. D, the central dogma of molecular biology states that 'DNA encodes RNA and RNA encodes proteins. Thus, information flows in one direction when genes are expressed: from DNA to RNA to protein.

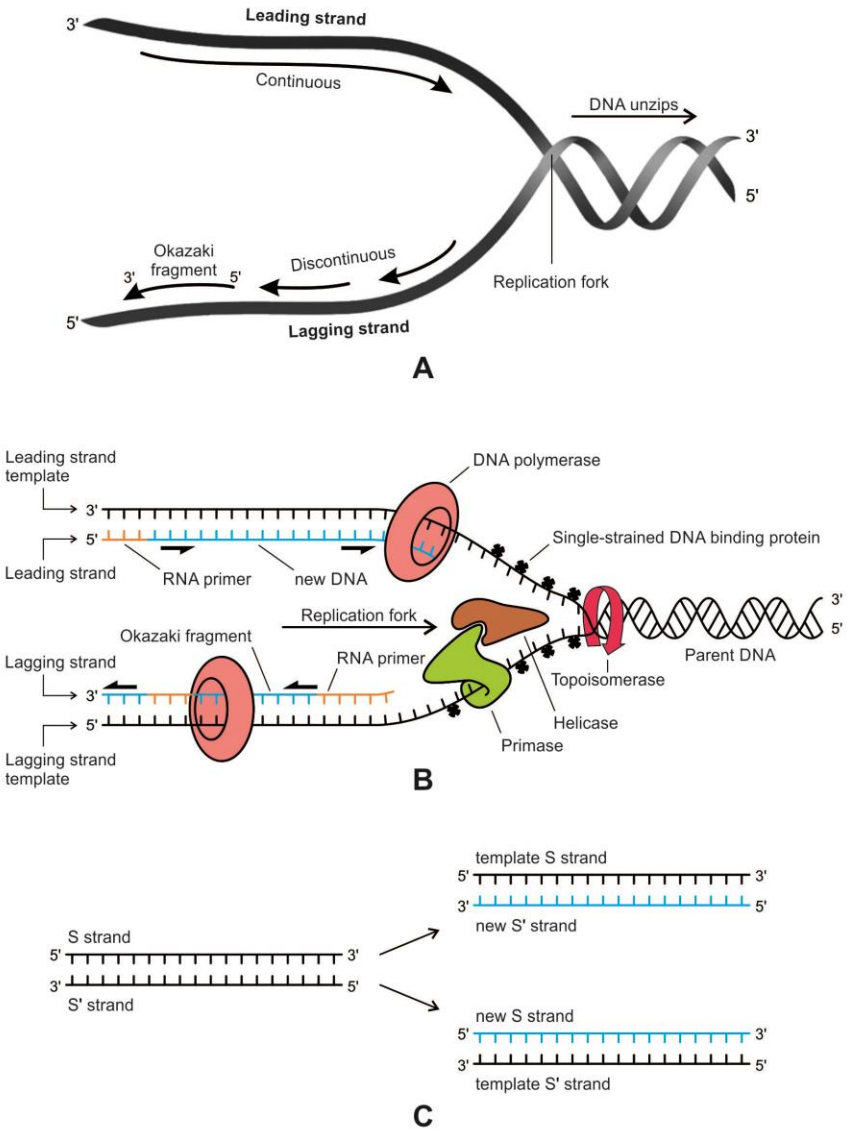


Figure 15. Replication of DNA by DNA polymerase. A, to begin DNA replication, DNA helicase enzyme causes the two parent DNA strands to unwind and separate from one another to form Y-shaped replication fork. Both new strands are synthesized in the 5´-to-3´ direction. The leading strand grows continuously forward, but the lagging strand grows in short discontinuous stretches called Okazaki fragments. B, many core enzymes—helicase, primase, and DNA polymerases collaborate at the replication fork and are involved in DNA replication. C, primordial DNA replication in protocell; Here, the short nucleobases of DNA were uncoiled and completely separated into two strands as S and S´, S can serve as a template for making a new strand S', while strand S´ can serve as template for making a new strand S (Fig. 14C). In this way double-helical DNA can be copied precisely.

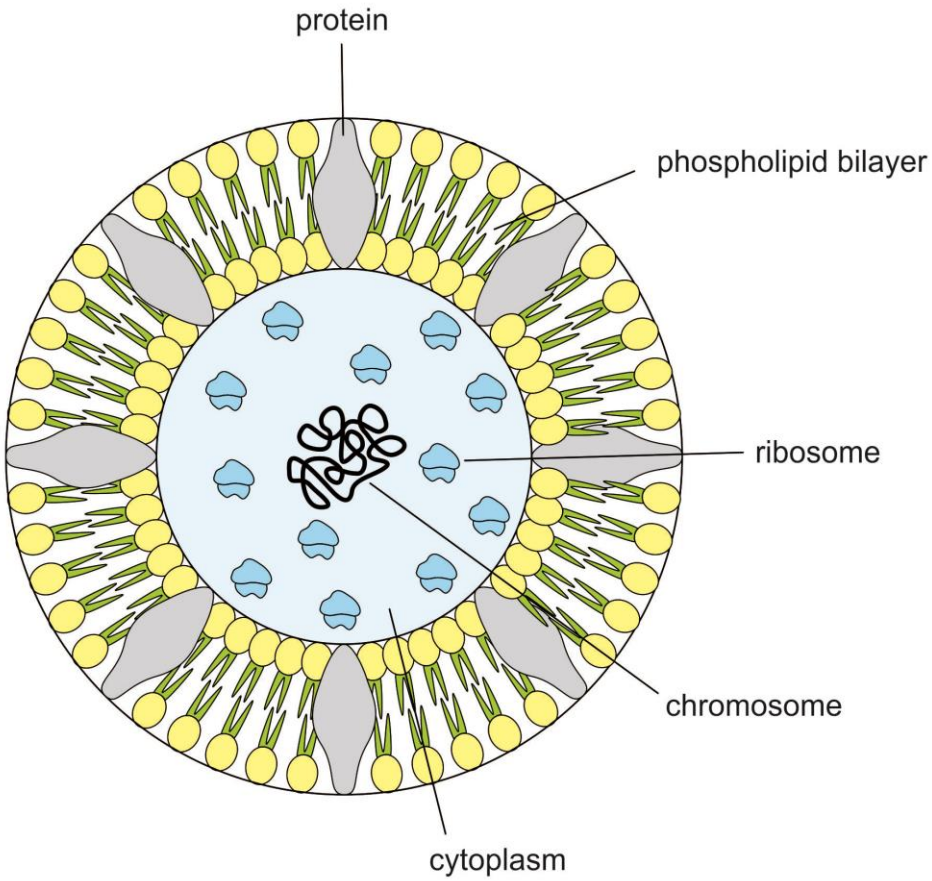


Figure 16. Reconstruction of the hypothetical first cell. The plasma membrane encloses a primitive cytoplasm that contains a prominent circular chromosome at the center; outside the chromosome, the most prominent structure was numerous ribosomes; other molecules like RNAs, various enzymes and proteins, and other translation machineries were too small to show.

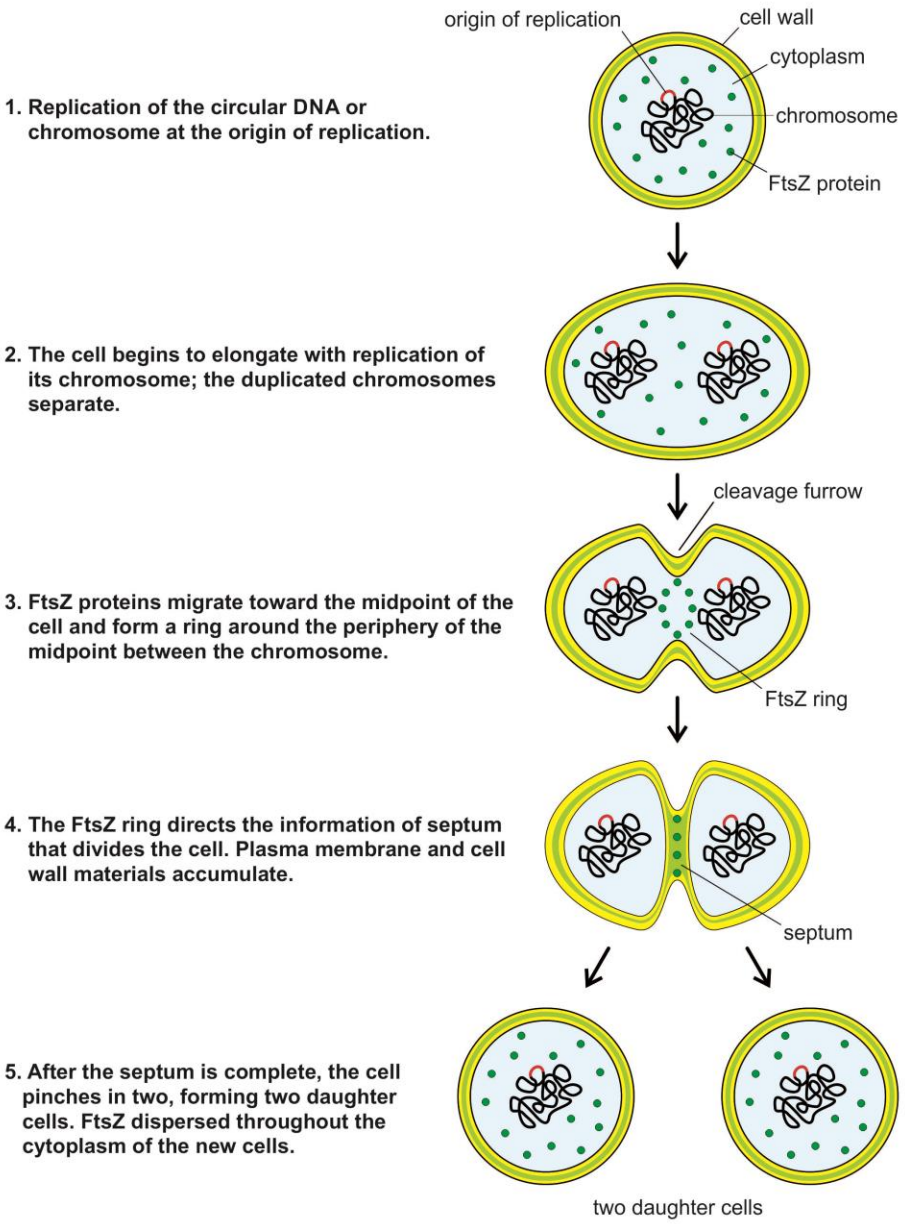


Figure 17. A hypothetical scheme of cell division in primitive first cell using modern bacterium as a guide. The cell-division cycle takes place in a cell leading to duplication of its DNA. The early cell must coordinate its growth, division, cell volume and shape with inheritance of genome. During the process, thousands of FtsZ molecules come together in the middle of the cell and form a circle like structure known as the Z-Ring. Z-rings are produced at the middle of the cell division leading to constriction. The replicated chromosome and cytoplasm separate into two new identical daughter cells.

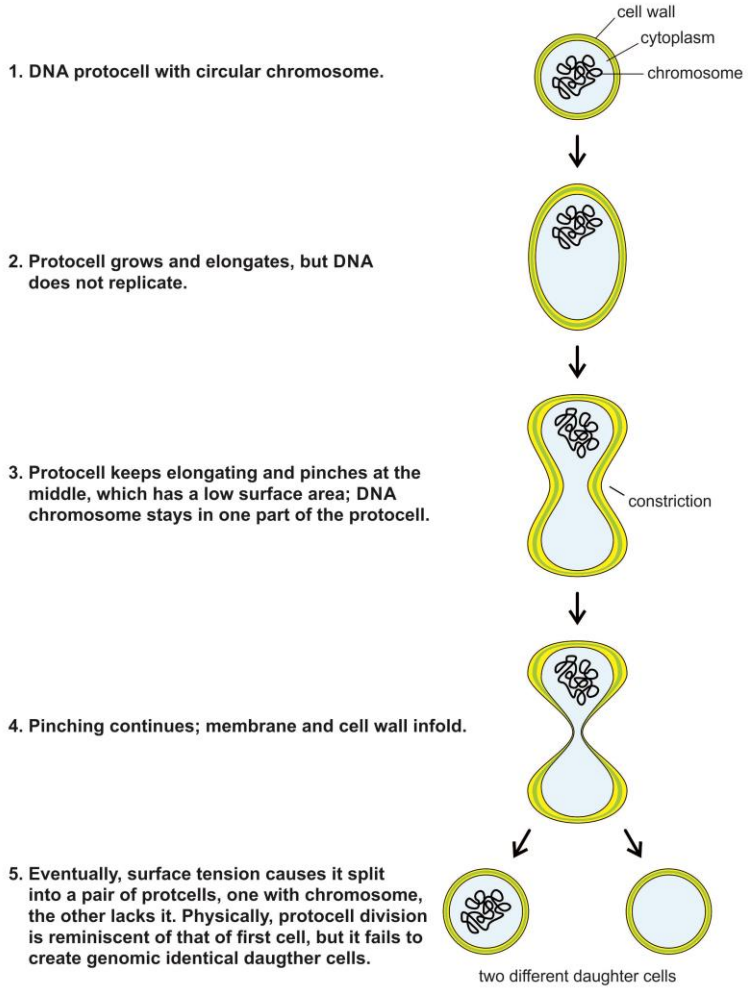


Figure 18. DNA replication was the major driver of symmetrical binary fission in early cells. Most likely there were many trials and errors before the cell division was perfected. Here a hypothetical scheme is shown where chromosome duplication was not perfected, thus giving rise to different kinds of daughter cells.

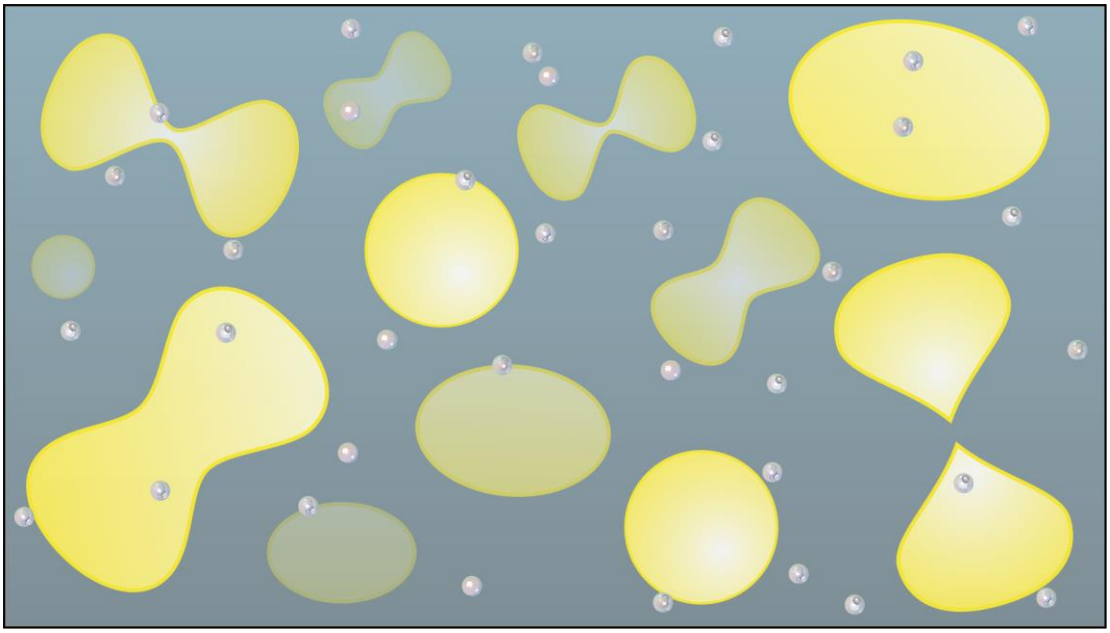


Figure 19. Once the cell division was perfected, the hydrothermal vent environment was crowded with a new generation of daughter cells.

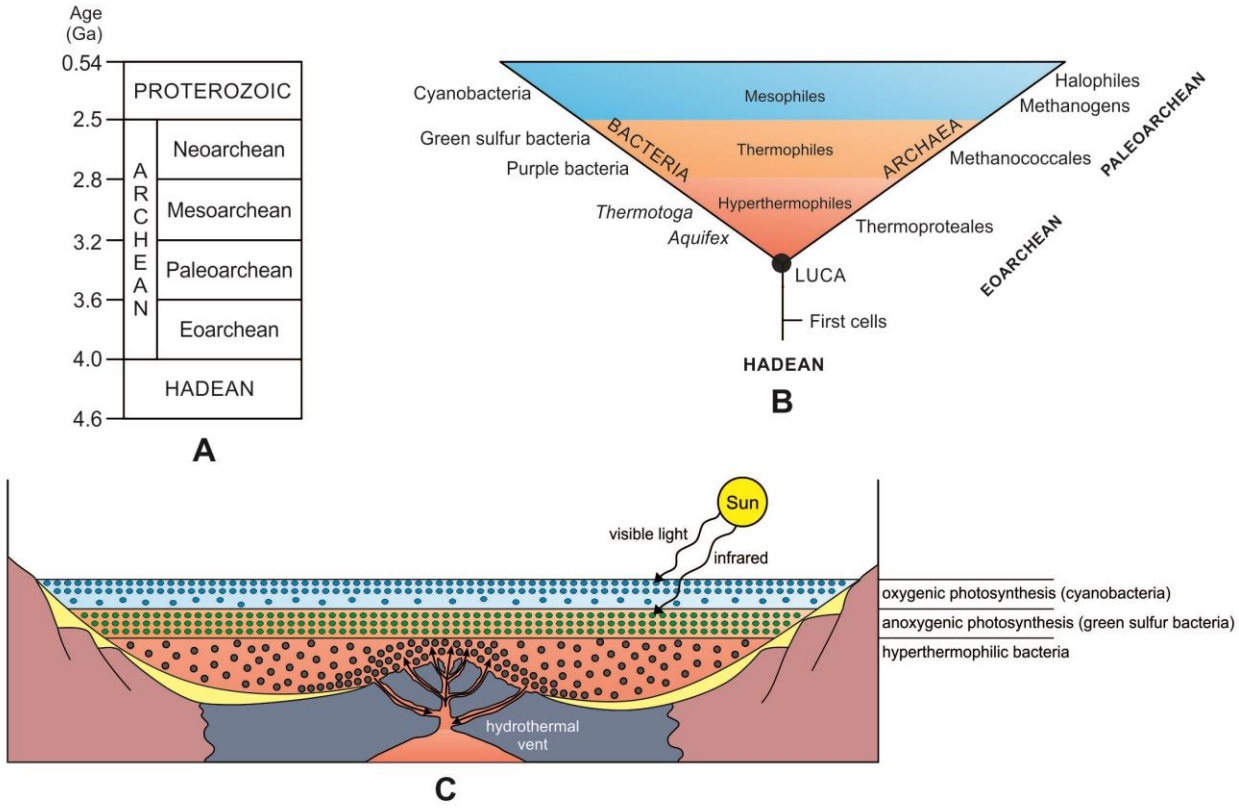


Fig. 20. The evolution of the Archean biosphere. A, delineates the geological timescale of the Archean eon. B, illustrates the origin and early evolution of life in an Archean hyperthermophilic, benthic crater-lake. Two distinct domains of life, Bacteria and Archaea, are preserved in the fossil record (~3.5 Ga) of the ancient Vaalbara continent. Both domains show the gradual reduction of thermotolerance over time: from hyperthermophilic to thermophilic to mesophilic. In the first hyperthermophilic habitat both bacteria and archaea appear. In the second, thermophilic habitat bacteria evolve as anoxygenic photosynthesizers. In the final mesophilic habitat, bacteria congregate at the upper surface of the crater-lake; they begin to tap solar energy, and evolve into oxygenic cyanobacteria. Over time, these cyanobacteria spread globally through the ocean and begin to produce oxygen. Hyperthermophilic archaea, on the other hand, evolve through two stages: as thermophilic Methanococcales, and then as mesophilic Methanogens and Halophiles, the latter thriving in the hypersaline environment of ponds and lakes. B, a cross-section of a hydrothermal crater-lake showing the evolution of photosynthesis. This occurs through three stages of an evolving microbial community: first, on the bottom, hyperthermophilic bacteria emerge; next, in the thermophilic middle stage, the anoxygenic photosynthetic green sulfur bacteria appear; in the final stage, at the upper mesophilic level, oxygenic photosynthesizing cyanobacteria form and begin the production of oxygen.