A Unifying Scenario on the Origin and Evolution of Cellular and Viral Domains

Claudiu I. Bandea

National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333

cbandea@cdc.gov / October 20, 2009

The cellular theory on the nature of life has been one of the first major advancements in biology. Viruses, however, are the most abundant life forms, and their exclusion from mainstream biology and the Tree of Life (TOL) is a major paradox in biology. This article presents a broad, unifying scenario on the origin and evolution of cellular and viral domains that challenges the conventional views about the history of life and supports a TOL that includes viruses. Co-evolution of viruses and their host cells has led to some of the most remarkable developments and transitions in the evolution of life, including the origin of non-coding DNA as a genomic protective device against viral insertion damage. However, one of the major fundamental evolutionary developments driven by viruses was probably the origin of cellular domains - Bacteria, Archaea and Eukarya - from the Last Universal Common Ancestor (LUCA) lineage, by evolving anti-fusion mechanisms. Consistent with a novel fusion/fission model for the population mode of evolution of LUCA, this paper presents a "cell-like world" model for the origin of life. According to this model the evolution of coupled replication, transcription and translation system (RT&T) occurred within non-living cell-like compartments (CCs). In this model, the ancestral ribosome originated as template-based RNA synthesizing machinery. The origin of the cellular genome as a centralized unit for storage and replication of genetic information within the CCs facilitated the evolution of the ancestral ribosome into a powerful translation machinery - the modern ribosome. After several hundred millions of years of providing an enclosed environment and fusion/fission based exchanges necessary for the population mode of evolution of the basic metabolism and the RT&T, the CCs evolved into the first living entities on earth - the LUCA lineage. The paper concludes with a proposal for a TOL that integrates the co-evolution of cellular and viral domains. This is one of a series of three articles that present a unifying scenario on the origin and evolution of viral and cellular domains, including the origin of life, which has significant t bio-medical implications and could lead to a significant paradigm shift in biology.

Introduction

The cellular theory on the nature of life has been one of the first major advancements in biology. The challenge to this theory was the discovery of viruses as infectious "filterable agents" at the turn of the last century [reviewed in (1-3)]. After several decades of research and debate about the nature of viruses that focused on the structural and biochemical properties of the transmissible infectious forms in the viral life cycle - the viral particles,- the dogma of viruses as viral particles became fully established [reviewed in (4)]. This dogma set viruses outside the mainstream biological and evolutionary paradigms.

It is becoming evident, however, that viruses are the most abundant life forms on Earth and that the repertoire of viral genes is greater than that of cellular genes (5-7). Moreover, it appears that viruses have played a major role in driving cellular evolution (4;8-13). Yet, viruses are not included in the Tree of Life (TOL), which represents exclusively the evolution of cellular domains - Archaea, Bacteria and Eukarya (14-16). It is difficult to put this major paradox in biology in a scientific perspective, or reconcile it with the current knowledge about viruses and cells (4;8;17;18).

As previously discussed, viruses have been historically identified with their viral particles (4;17), which are highly specialized structures used for transmission of viruses to new host cells [for a general review of this and other facts about viruses see (3)]. Based on an alternative view about the nature of viruses, in which viruses are defined primarily based on their properties during the intracellular stage of their life cycle, when the viral molecules are more or less "free," or dispersed within their particular environment, the host cell, I suggested the concept of *molecular structure* and labeled viruses as *molecular organisms* (4). Supporting

these new concepts is a hypothetical evolutionary model for the origin of ancestral viruses from parasitic cellular species that in order to take full advantage of the host resources fused with their host cells (4). The concept of a virus as molecular organism sets up the foundation for including diverse viral lineages, currently referred to as "plasmids," "endogenous viruses," "transposable elements," "viroids," "phages," and "viruses" within the same domain of biological organization - the viral domain.

Because viruses are parasitic or symbiotic intracellular organisms, their origin and evolution have been intrinsically linked to their host cells [reviewed in (4;11;17)]. When viruses were first identified as "filterable agents" at the turn of the last century, it was thought that they were minute parasitic cellular species that originated from more complex intracellular parasites by reductive evolution. After it was shown that the viral particles of some viruses have a non-cellular structure and a simple composition, reminiscent of those of the hypothetical first living entities, the hypothesis on their ancient origin from pre-cellular organisms became prevalent. Later on, when the mobile genetic elements were discovered, the endogenous theory on the origin of viruses from escaped cellular genetic material became popular. However, very few comprehensive scenarios for the evolutionary origin of viruses and co-evolutionary events between viruses and their host cells have been proposed [reviewed in (4)]. On the other hand, the origin and evolution of cells has been a very active and debated evolutionary issue addressed in hundreds of publications [reviewed in (15;19-28)]. The hypotheses on the origin and evolution of cellular domains fall within two major lines of thought. The traditional view is that the first cellular organisms were Bacteria from which Archaea and Eukarya eventually evolved, but the intriguing hypothesis that the eukaryal and archaeal lineages originated first is gaining popularity [reviewed in (15;20-28)].

This article presents a broad evolutionary framework that integrates the co-evolution of cellular and viral domains. To my knowledge, this is the first attempt to develop a unifying scenario on the evolution of cellular and viral domains, including the origin of life. Although speculative, many of the ideas, models and hypotheses associated with this unifying scenario are open to immediate experimental and analytical evaluation. The intent of this scenario is to stimulate new interpretations and hypotheses, a process that is essential for productively integrating the large amounts of experimental data and knowledge that are currently being generated at an overwhelming rate; unfortunately, the development of concepts and ideas have not kept up with the extraordinary advances in experimental biology.

This paper begins by exploring the co-evolution of viruses and their cellular hosts, which has led to some of the most significant fundamental transitions in the history of life, including the evolution of major cellular genomic features, such as non-coding DNA (ncDNA) and spliceosomal introns. However, the major role played by viruses was probably in the origin and evolution of the three cellular domains - Bacteria, Archaea and Eukarya - from the "Last Universal Common Ancestor" (LUCA) lineage by developing anti-fusion mechanisms. A novel integrative "cell-like world" model for the evolutionary origin of LUCA lineages expands this unifying scenario on the evolution of cellular and viral domains to include the origin of life. According to this novel model for the origin of life, the evolution of coupled replication, transcription and translation system occurred in non-living, cell-like compartments and it started with the origin of the ancestral ribosome as template-based RNA replicating machinery. The evolutionary diversification of the ancestral ribosome led to the origin of the cellular genome as a centralized unit for storage and replication of genetic information within cell-like compartments, and to the origin of the modern ribosome as a powerful translation machinery. The paper concludes with a proposal for a new TOL, which integrates the co-evolution of cellular and viral domains. This is one of a series of three papers that present a broad, unifying scenario on the origin and evolution of viral and cellular domains (4;29).

Co-evolution of cellular and viral domains

Because of their molecular structure, viruses evolved highly diversified parasitic and symbiotic life styles and unique molecular mechanisms that would not have been feasible for species maintaining a cellular structure within the host cell. In an adaptive response, the hosts' evolutionary changes were equally remarkable (4). To explore this co-evolution, it helps to think about the origin and evolution of viruses in context of two distinct evolutionary periods – before, and after the origin of the three cellular domains - Archaea, Bacteria and Eukarya (4;11). This section of the paper addresses this more recent evolutionary period.

Along with other parasites and symbionts, viruses have played a critical role in the evolution of their hosts driven by parasitism, disease, and symbiosis. Many aspects of this intricate co-evolution under classical Darwinian selective forces has been discussed at large in other publications (13;30-33). In addition to these selective forces, viruses have shaped the evolution of their hosts by well known direct genetic mechanisms, such as lateral gene transfer (LGT), recombination events, and modulation of the host's gene expression [reviewed in (13;30-33)]. Because of the ubiquitous presence and universal distribution of viral organisms among all cellular species throughout the history of life, their contributions to the evolution of their hosts have been profound and highly diverse. For instance, there is strong evidence that in humans and other mammals viruses have played a direct role in the evolution of placenta, immune system, and the nervous system [reviewed in (13;33;34)]. And, this might be just the tip of the iceberg.

Here I outline three novel hypothetical co-evolutionary

models between viruses and their hosts that, if true, would have a substantial impact on our current understanding of: (a) speciation in Bacteria, Archaea, and eukaryal asexually reproducing organisms, (b) evolution of major histocompatibilty complex (MHC) in vertebrates, and (c) evolution and function of the "junk" or non-coding DNA (ncDNA) in Eukarya. These examples seem to be completely unrelated subject matters. However, these and many others seemingly unrelated biological phenomena are part of a common theme: the intricate co-evolution of viruses with their hosts.

A virus-induced speciation mechanism in asexually reproducing cellular lineages

The species concept in Bacteria, Archaea and asexually reproducing eukaryal organisms is not well defined and the speciation mechanisms poorly understood (reviewed in (35-37). Based on a novel endosymbiotic evolutionary model between viruses and their host, it is argued here, that many archaeal, bacterial and some eukaryal lineages, including many human bacterial pathogens and parasites, have evolved as virus-induced "cell-lines," or viral clones. By providing their host with a beneficial phenotype, such as antibiotic resistance, an exogenous virus that produces viral particles for horizontal transmission has the opportunity to evolve into a vertically-transmitted endogenous, or plasmid-like viral organism by inducing successful clonal expansion of the host cell they infect.

By replacing all non-infected related individuals in a particular niche, these virus-induced host clonal lineages become dominant genotypes. Overtime, the members of this new species evolve by divergent evolution into a genetically less homogenous population only to have another viral species in the future establish a symbiotic relationship with a particular individual of this population leading to its clonal expansion and a new speciation event. Possibly, this microbial expansion\replacement wave induced by symbiosis between co-evolving viral organisms and the hosts has been a major speciation mechanism in Bacteria, Archaea, and asexual eukaryal organisms.

The ubiquitous presence of endogenous and plasmid-like organisms and their ancestors, the phages, in Bacteria and Archaea supports this model for the origin and evolution of diverse lineages. The full implication of this hypothetical model for understanding diversification and speciation phenomenon in Bacteria, Archaea and Eukarya, which include important human pathogens and parasites, remains to be tested and evaluated in laboratory and theoretical modeling.

Evolution of MHC allelic diversity and anti-MHC allogeneic cytotoxic response

The astounding allelic diversity of MHC in many vertebrates, particularly in humans who have several hundred alleles, and the strong anti-MHC allogeneic cytotoxic response (i.e. transplantation antigens response) are among the most debated and enigmatic evolutionary issues in immunology (38-42). The following evolutionary model offers a plausible evolutionary explanation for these phenomena that might help with the understanding of the epidemiology and the dynamics of viral transmission and could lead to new approaches in vaccine development.

Many viruses, such as retroviruses, use modified host cellular membranes as envelopes for the assembly of their viral particles. For example, the envelope of human immunodeficiency virus (HIV) contains viral glycoprotein complexes composed of the viral envelope proteins gp41 and gp120, which enables the viral particles to attach to and fuse with their target cells (43). In addition to these viral glycoprotein complexes (approximately 70 per HIV viral particle), the viral envelope contains an excess of host-derived proteins, including MHC antigens which are up regulated in the HIV infected cells (44;45). It is likely that during the co-evolution of viruses and their vertebrate hosts, some viral species evolved mechanisms for including host self-antigens, such as MHC, on the surface of their viral particles in order to mask the viral antigens from immune recognition during their transmission to new host cells within the same individual host and, more importantly, during their transmission to new members of the population. In an adaptive evolutionary response, the vertebrate species diversified their MHC alleles and increased their cytotoxic response to allogeneic MHC antigens.

According to this model, viral particles that carry MHC proteins on their surface have a better chance to escape host immune recognition and initiate a productive infection in hosts that have identical or similar MHC as compared to infecting hosts that have allogeneic MHC. In a syngeneic match, these viral particles are likely to be regarded as "self-antigens," whereas in an allogeneic match they would be recognized as "foreign-antigens" and possibly neutralized before they have a chance to start a productive infection. Hypothetically, the transmission of viral infections, such as HIV, between individuals with identical or similar MHC alleles occurs at a higher rate than between individuals with different MHC alleles, generating the selecting force leading an increase in MHC allelic diversity and anti-MHC allogeneic direct cytotoxic response.

This evolutionary model can be evaluated epidemiologically by scoring for natural transmission of viruses between syngeneic and allogenic MHC individuals in both, humans and animal populations. However, it is relatively straightforward to test this hypothesis experimentally in animal models with known MHC genotypes. If proven correct, this model has significant implications. A potential practical application is a vaccine development approach that would target the host antigens on the surface of enveloped viruses rather than the viral antigens. Interestingly, almost two decades ago, it was fortuitously revealed that a control group of macaques immunized with uninfected human cells were protected from challenge with simian immunodeficiency virus (SIV) grown in human cells expressing the same MHC alleles

(46;47). The present model offers an evolutionary and conceptual rationale for implementing such a vaccine strategy against some of the most devastating viral infections, such as HIV.

An anti-viral protective role for eukaryal ncDNA, including introns

As described next, viruses have also played a significant role in the evolution of the basic organization and architecture of the host genome. The "junk," or non-coding DNA (ncDNA), including the spliceosomal introns, is one of the most prominent and puzzling eukaryal genomic feature [reviewed in (48-58)]. In humans, for example, it is estimated that more than 98% of the genome consists of ncDNA, with about a quarter of it in introns (59). Although, several hypotheses have been advanced on potential functions of ncDNA (48;60;61), the mainstream view remains that ncDNA is selfish or parasitic in nature (62;63).

Two decades ago, it was proposed that ncDNA evolved as a genomic protective mechanism against viral insertion mutagenesis by serving as a sink for the integration of proviral genomes and other endogenous mobile viral genetic elements (64). In agreement with this evolutionary scenario, it is hypothesized here that gene splicing evolved to allow for the presence of ncDNA within transcribed regions, which are preferred targets for the integration of viral genomes (65-68). It is well known that much of the ncDNA is composed of remnants of viral sequences, and that the eukaryal introns resemble the group II self-splicing introns, which in turn resemble retroviral elements. Likely, these elements are evolutionary related (51;57) and it is highly probable that the spliceosomal machinery originated from symbiotic endogenous viral species that coevolved with their host to protect the coding regions from insertional damage (more on the selective forces leading to the evolutionary origin of introns and spliceosomal machinery in the next section).

In the absence of an obvious function, the initial classification of ncDNA as parasitic or "selfish DNA" made sense (62;63). However, it is puzzling that Bacteria and Archaea, which host numerous lysogenic viruses and transposable elements, do not contain much ncDNA. As previously discussed (64), these organisms evolved alternative protective mechanisms against viral insertion mutagenesis, such as specific genomic sites for integration. The evolution of these mechanisms in Bacteria and Archaea is strong evidence that the selection pressure for developing protective mechanisms against viral insertion mutagenesis in these organisms is very strong. However, this selection pressure was probably stronger in eukaryal multi-cellular species, in which the viral insertion events could potentially induce cancer (see below). And yet, no protective mechanisms against viral insertion mutagenesis have been found in eukaryal species, which is puzzling.

The current evidence indicate that similar to bacterial and archaeal cells, the eukaryal cells have evolved mechanisms for removal of ncDNA (69). The fact that very large quantities of ncDNA are present in many eukaryal species, however, suggests a relatively strong selection for its accumulation. As mentioned above, defense mechanisms against endogenous viral integration were especially critical in multi-cellular organisms, in which the modulation of the host's gene expression by the integrating viral sequences could lead to cancer [reviewed in (70-73)]. Considering the extent of somatic cells turnover during the reproductive life span of multi-cellular organisms, in the absence of a protective mechanism, the number of integration events in these somatic cells potentially leading to cancer would be enormous, which would reduce the chances for evolutionary survival of these species. Therefore, it would be expected that these species evolved protective mechanisms and it is likely that one of these mechanisms was the accumulation of ncDNA.

The evolution of alternative protective mechanisms in the form of specific integration sites by organisms that have little or no ncDNA, as well as the evolution of other antiviral mechanisms across all cellular domains (74-78), is strong circumstantial evidence for a strong selective pressure to evolve protective mechanisms against viral insertion mutagenesis in eukaryal cells, such as ncDNA. However, the protective function of ncDNA has undeniable direct statistical support: the more ncDNA the less chance for insertion damage of the coding regions. For example, in humans, in which more than 95% of the genome is ncDNA, likely, the ncDNA reduces insertion mutagenesis by an equivalent percentage, or possibly more (see below). Early on, during the evolution of this protective mechanism, when there was little ncDNA, there were probably enhancing mechanisms such as recombination (79), which could have targeted the integration of the viral genomes in host genomic regions containing homologous viral ncDNA sequences. It is highly probable, therefore, that some of the recombination machineries evolved at least partially to facilitate the insertion of invading viral genomes within the host chromosomal regions containing ncDNA, which enhanced the protective role of ncDNA.

The protective function of ncDNA against cancer, as well as the potential role of recombination as an enhancing mechanism could be easily tested. For example, transgenic mice carrying DNA sequences homologous to infectious retro-viruses, such as murine leukemia viruses (MuLV), might be more resistant to cancer induced by experimental MuLV infections as compared to controls. Although, additional supporting evidence should help with understanding the full spectrum and relevance of the protective function of ncDNA, it should be recognized that the strong statistical support for this function (see above) is a fact.

As discussed in a different paper in this series (29), in addition to using viral DNA as a protective mechanism against viral insertion mutagenesis, the hosts and their endogenous viruses have coevolved elaborate, viral protein-based protective mechanisms against pathogenic viruses. This phenomenon points once more to co-evolutionary

versatility between viruses and their hosts made possible by the viral molecular structure (4). As described next, however, this structure has probably played an even more remarkable role in the evolutionary origin of the three cellular domains and their ancestor, the Last Universal Common Ancestor (LUCA).

A fusion/anti-fusion model for the origin of cellular and viral domains

The foundation for modern thinking on the evolutionary origin Archaea, Bacteria, and Eukarya is the modern TOL, which is based on the phylogenetic analysis of conserved cellular genes (14;15). One of the most difficult, unresolved problems in evolutionary biology is the universal evolution of basic cellular features leading to the origin of LUCA and the three cellular domains (80). To solve this intriguing problem, which has yet to be satisfactorily explained in models involving natural selection, Carl Woese proposed the concept of communal, or collective evolution of pre- and early cellular entities driven by LGT (19;80;81). Not surprising, by circumventing natural selection as the driving force for the universal evolution of the basic cellular features, this highly relevant concept was questioned (82). The evolutionary model presented next proposes an alternative mechanism to LGT for collective evolution. In this model, the mechanism and process of collective evolution were fully driven by natural selection.

Based on a suggestion about the potential role of cellular hybridization, or fusion, in the early evolution of cells and the origin of ancestral viruses (17), it is proposed here that the LUCA lineage evolved in a population mode similar to that of sexually reproducing species. In this mode of evolution, the individual genotypes are ephemeral and what persists and evolves is a gene pool (83). In sexually reproducing species, the mechanism for their population mode of evolution is sexual reproduction. By analogy, it is proposed here that the life cycle of LUCA lineage was based on fusion/fission events (i.e. primitive sex), which led to the collective, universal evolution of basic cellular mechanisms, including the coupled replication, transcription and translation system (RT&T). The significance of fusion/fission events in the evolution of pre-cellular and early cellular entities has been also discussed by Otto Kandler and Gunter Wachtershauser, but from a different evolutionary perspective (22;84-86).

In the population mode of evolution presented here, the members of the LUCA lineage were genuine cellular organisms that followed a life cycle in which two or more parental cells fused and reproduced by producing daughter cells by a fission-like mechanism. Eventually, some of the fusing members in the LUCA population developed a dominant (i.e. parasitic) life cycle by fusing with other (i.e. host-like) members of the LUCA lineage and producing preferentially their own progenies. This scenario resembles the fusion model on the origin of viruses as molecular organisms from parasitic cellular species that fused with their host cells (4). These early parasitic molecular

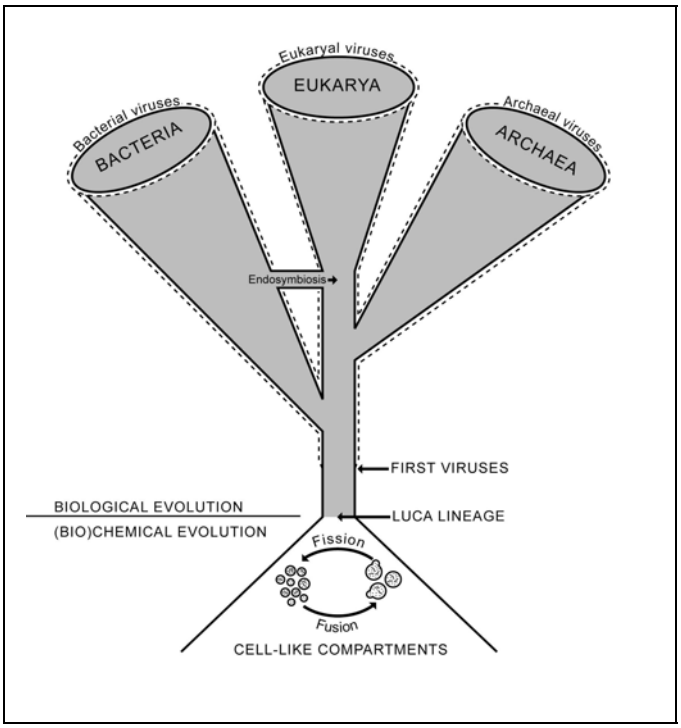

Figure 1. TOL showing the origin and evolution of cellular and viral domains.

organisms were probably the first members of the LUCA population to undergo reproductive isolation and speciation,leading to the origin of first viral lineages (Fig. 1). Likely, these viral species played a significant role in the eventual origin of the three cellular domains - Bacteria, Archaea, and Eukarya - which branched off the LUCA lineage by developing anti-fusion mechanisms, leading to their reproductive isolation and speciation. The earliest cellular species that branched off the LUCA lineage were probably those that developed anti-fusion mechanisms consisting of murein-based cell walls, which led to the evolution of Bacteria. A second major anti-fusion series of evolutionary events within the LUCA lineage followed, leading to the origin of Archaea; the anti-fusion mechanism in Archaea was associated with changes in the cellular membrane biochemistry, such as lipid chirality and composition [reviewed in (80;87-91)]. As described next, the origin of Eukarya from the post-Bacteria/Archaea LUCA lineage was a more involved process [reviewed in (25;92-94)].

In the present scenario on the evolution of the early cellular species, the fusion/fission life cycle of the LUCA lineage was the main mechanism and selective force behind the early evolution of the cellular cytoskeleton and phagocytosis as well as the evolution of the cyto-membrane system. The dynamic process of complete or partial fusion events of two or more parental cells integrating their metabolism, replicating their multi-copy genome and generating daughter cells by fission or internal morphogenesis of progenies, drove these intra-cellular

innovations. The structural and functional compartmentalization of the ever growing assembly of fusing cells was inevitable as an evolutionary opportunity, and it was achieved by a mechanism in which some of the engulfed parental cells maintained their cellular membrane and structural identity at least temporarily during their life cycle [these cellular assemblies could be envisioned as "chimeric organisms," with the annotation that as members of the LUCA lineage the parental cells were evolutionarily highly related, a situation similar to that among individuals in sexually reproducing species]. Considering the potential dynamic participation of viral species in this process, as envisioned also by Patrick Forterre in a different evolutionary context (95), the amplitude of the selective forces leading to the compartmentalization of the cell and the eventual evolution of the modern eukaryal type of cellular organization skyrocketed.

In the present scenario on the evolution of the LUCA lineage, the RT&T and the overall cellular features of first Bacteria and Archaea were similar to those of the LUCA population at the time they branched off. Likely, some eukaryal-like cellular features such as an ancestral cytoskeleton and cyto-membrane system evolved relatively early in the LUCA lineage and, hence, were present in the ancestral Bacteria and Archaea. However, to accommodate new metabolic features and life cycles that were imposed by the anti-fusion mechanisms, the bacterial and archaeal organisms eventually lost some of these features. Due to their apparent structural and functional simplicity, Bacteria and Archaea are usually perceived as less "evolved" than Eukarya. On the contrary, if evolution could be quantified, based on population size, generation time, and length of evolution, Bacteria and Archaea are more "evolved" than Eukarya by an enormous factor.

Unlike Bacteria and Archaea, the members of the LUCA lineage continued their fusion/fission life cycle that drove the evolution of many of the modern eukaryal cellular features, including mitosis and meiosis, which led to the evolution of true sexual reproduction. It would be difficult to envision the evolution of these eukaryal features in cellular lineages that did not followed a fusion/fission-based life cycle. Of prominent significance to the evolution of Eukarya has been the origin of the nucleus, mitochondria, and plastides. In modeling the evolutionary origin of the ancestral eukaryal organisms, however, it is critical to co-address the origin and evolution of other eukaryal features, such as the cytoskeleton, the cyto-membrane system (whose basic activities in modern eukaryal cells are still based on fusion/fission events), and other organelles, including the peroxisomes and hydrogenosomes (92;96).

The hypothesis for the endo-symbiotic origin of mitochondria from a non-fusing ancestral alpha-proteobacteria is well supported (97). Intriguingly, the majority of the mitochondrial proteins are coded by nuclear genes. The current hypotheses account for this phenomenon by suggesting that these genes were transferred from the mitochondrial genome to the host genome through LGT. It is likely, however, that at the time an ancestral alpha-proteobacteria evolved as an endosymbiont, they shared many genes with the host (*i.e.* the post-Bacteria/Archea LUCA lineage), and it was only a matter of time until some of the duplicated genes in the mitochondria or other endosymbionts were lost. This evolutionary scenario might be valid also for other cellular organelles, such as the peroxisomes, which as suggested by Christian de Duve probably lost all their genes, making their endosymbiotic origin less apparent (92). If this scenario is correct, it is fascinating that some endosymbiotic organisms lost their intrinsic genome while maintaining their cell-like structure, whereas many viral species lost their cellular structure within the host cell, but maintained their genome.

The dispersion of the nuclear membrane and nucleus during the division cycle in most eukaryal cells and the structural continuity of the nuclear/endoplasmic membrane system are features congruent with a model for the origin of the nucleus as described here. Likely, the nuclear membrane originated as a physical barrier to protect the host genome against insertion mutagenesis by invading fusing viral species. In fact, it is likely that, at least partially, the eukaryal membrane system evolved in order to compartmentalize the viral infection, allowing the infected host cells to survive. Expectedly, however, some viruses evolved counter mechanisms that allow them to infect these intracellular compartments, leading to a continuous fusion/anti-fusion arms race between viruses and host cell, which shaped the evolution of eukaryal cellular organization and the viral molecular structure. After some viral species evolved mechanisms for entering the nucleus, the hosts evolved other protective mechanisms against insertion damage, such as ncDNA and spliceosomal introns, which was discussed in the previous section. The evolution of gene splicing required a mechanism for decoupling transcription from translation, so that the translation would only proceed on spliced mRNAs (98). It is hypothesized here that the decoupling mechanism took advantage of the existence of nuclear membrane. Certainly, in the absence of a nuclear membrane, the eukaryal cells could have evolved alternative decoupling mechanisms. However, nothing about splicing seems to indicate that the eukaryal cells have ever used alternative decoupling mechanisms, suggesting that spliceosomal introns and spliceosomes evolved after the origin of nuclear membrane and nucleus.

A relevant issue regarding the evolution of the three cellular domains is the potential "genetic bottleneck" associated with their origin. In the present model, due to the population mode of evolution of the LUCA lineage in which, basically, each member was genetically representative for the whole population, the potential for a "genetic bottleneck" associated with the origin of Bacteria and Archaea from the LUCA lineage was minimal. Basically, the first bacterial and archaeal species were genetically similar to the LUCA population at the time of their origin. Interestingly, if the LUCA lineage was open to

the evolution of bacterial- or archaeal-like anti-fusion mechanisms over long periods, then numerous new bacterial and archaeal lineages probably originated independently from the evolving LUCA lineage within those open periods, suggesting a polyphyletic origin of these domains.[1]

The evolution of post-Bacteria/Archaea LUCA lineage that eventually led to the origin of modern Eukarya continued in a population-wide mode of evolution driven by its fusion/fission life cycle. However, this lineage eventually diversified through anti-fusion mechanism, which led to evolution of some of the ancestral eukaryal asexual lineages, or through selective fusion mechanisms, led to true sexual lineages speciation events, giving rise to the ancestors of some of the extant sexually reproducing single-cell eukaryal lineages. Probably, the most spectacular "anti-fusion" event in eukaryal speciation was the evolution of multi-cellular species. The origin of these species from sexually reproducing single-cell organisms followed an evolutionary path in which some of the daughter cells remained physically and functionally linked in a "colony" in order to protect the parental fusing cells (which eventually become the "germ-line") from nonspecific fusion events (including viral fusions) in their quest to fuse with similar cells from related "colonies" on the road of sexual speciation in multi-cellular organisms.

The evolutionary fate of the post-Bacteria/Archea/Eukarya LUCA lineage (i.e. eukaryal-like organisms that did not acquire the proteobacteria or other non-fusing endo-symbionts) is enigmatic. Possibly some of the last species branching off this remarkable lineage before extinction were the ancestors of some of the viral lineages that infected the early eukaryal species. Whatever the fate of the LUCA lineage, its population mode of evolution driven by a fusion/fission life cycle set the stage for the origin of viral and cellular domains. However, as discussed in a different paper (4) in this series, it is likely that many of the extant viral lineages originated more recently from bacterial, archaeal, or eukaryal parasitic species by fusion, or fusion-like mechanisms.

In the unifying evolutionary scenario the origin of the viral and cellular domains presented here, the early LUCA lineage was a genuine cellular organism with a well developed metabolism and RT&T, and with a fusion/fission-based life cycle. In line with this scenario, it is postulated here that LUCA originated from non-living, cell-like entities that followed a fusion /fission-based "life cycle" analogous to that of the LUCA lineage. This extraordinary transition period from chemical to biological evolution is described next in a "cell-like world" model for the origin of life.

A "cell-like world" model for the origin of life

When Darwin and Wallace proposed more than a century and a half ago that natural selection drives the evolution of all life forms on Earth, the origin of life entered the realm of scientific exploration. This exploration, however, has been strongly influenced by diverse philosophical views about the nature of the first living entities. Equating the origin of life with self-replicating molecules, such as RNA, with self-sustaining chemical reactions or metabolic pathways, with virus particle-like or other types of non-cellular entities, or with cell-like entities, each impose different conceptual and experimental approaches. It is not surprisingly, therefore, that the views about the origin of life and the experimental approaches in this field have been extremely diverse [reviewed in (20;28;89;99-110)].

During the last century, two major conceptual platforms have dominated the thoughts on the origin of life. One was based on the proposals and research promoted primarily by Oparin, Haldane, and Miller that life originated in a "primeval soup," or broth-like environment, in which chemical evolution led to the origin of the first living entities. The other platform, which was developed more recently, addressed the nature of the first living entities by proposing that they were self-replicating RNA molecules [referred to as "RNA world" (111)]; this view was based on the findings that RNA molecules can serve both as carriers of genetic information and as enzymes. It is becoming increasingly evident, however, that neither of these conceptual frameworks can fully integrate the current data and thinking in this field [discussed in (28;103;104;109;112)].

For reasons well rationalized, the transition from chemical to biological evolution occurred in a compartmentalized environment (20;28;103;110;112-116). To support the enormous number of dynamic interactions among these compartments, which was necessary for complex chemical evolution, it is proposed here that these compartments allowed inter-compartmental exchanges and, very importantly, this process occurred in a self-sustained mode. Based on these fundamental principles, it is likely that these dynamic compartments, which were probably bound by a lipid-based membrane and were associated with the surface or with the pores of inorganic formations or conglomerates in an aqueous environment, followed a "life cycle" similar to that of the LUCA lineage (see Fig.1):

(i) They assembled spontaneously;

(ii) They grew by fusing with other compartments;

[1] Assuming that the period for the origin of multiple lineages within each cellular domain did not overlap that of the other domains, this polyphyletic model is compatible with the topology of the current TOL. For example, if multiple bacterial lineages originated from the LUCA lineage independently over a period of one billion years in which the LUCA lineage evolved, followed by a similar period for the origin of multiple archaeal lineages, then the Bacteria and Archaea would appear as monophyletic groups on the TOL although they would have polyphyletic origin (see also a related discussion, later in this paper, about the mysterious canonical pattern in the evolution of highly conserved genes, such as ribosomal genes, across the three cellular domains).

(iii) They reproduced by splitting into smaller compartments that continued the fusion/fission cycle - a cycle that initially was supported by thermal, mechanical, or osmotic energy.

These self sustaining compartments developed in association with a surface or solution based auto- or hetero-trophic metabolism (99;100). It took these compartments, which are referred to here as cell-like compartments (CCs), several hundreds of millions years of providing an enclosed environment and fusion/fission based exchanges necessary for the population mode of evolution (i.e., collective evolution) of basic metabolism and RT&T to evolve into genuine cells, the LUCA lineage.

The origin of ancestral ribosomes as RNA synthesizing machineries

The origin of coupled replication, transcription and translation, the RT&T, is enigmatic (117-120). It is hypothesized here that, the evolution of RT&T started with the origin of the ancestral ribosome as RNA synthesizing machinery (Fig.2). The evolutionary diversification of the ancestral ribosome eventually led to the origin of the cellular genome and the evolution of the modern ribosome as a translation apparatus. Interestingly, the idea that the ancestral ribosome originated as a nucleic acid replicating machinery was proposed in the past, but for a different rationale; specifically, to explain the origin of the three-nucleotide-based genetic code by hypothesizing that the ribosome evolved from a "RNA triplicase," a hypothetical ancestral enzyme that presumably polymerized nascent RNA strands in three-nucleotide steps (121), [discussed also in (117)].

Among the products of early chemical evolution were amino acids and ribo-nucleotides (101-103;122;123). The amino acids had the potential property of being polymerized into polypeptide chains, or proteins with infinite structural and reactive properties. The ribo-nucleotides had the potential to be polymerized into RNA chains, generating linear information based on the order of monomers in the RNA chain. Fundamentally, the RT&T evolved as a system for storing, replicating, and transferring (amplifying) the linear information encoded in nucleic acids into the three-dimensional information encoded in proteins. The goal of the following model is to outline some of the critical steps in the evolution of this information processing system.

Unlike amino acids and fatty acids, which can be easily produced experimentally in simulated early Earth conditions, and are found in meteorites and, presumably, throughout the Universe (124), the synthesis of the ribo-nucleotides required a rather complex bio-chemical pathway (101;102;122;123). It is likely, therefore, that before the origin of RNA, a relatively advanced ancestral metabolism that was able to produce complex organic molecules, such as nucleotides, was already operating (107). As proposed below, these early metabolic reactions led to the synthesis of the first RNA molecules and eventually to the origin of ancestral ribosome as RNA synthesizing machinery within the CCs.

Very little is known about the early synthesis of the ribo-nucleotide monomers and the synthesis of the first RNA molecules (101;102;122;123). Thus, the scenarios about the "RNA world" usually start with ready-made, self-replicating RNA molecules. One way of thinking about the origin of self-replicating RNA molecules is that a random population of RNA molecules was somehow synthesized and some members of this population were able to self-replicate by chance, which jump-started the RNA evolution. However, a template-based RNA self-replication process is more complicated than it is usually perceived, because in addition to a relatively high concentration of ribo-nucleotide monomers, this process would require the synthesis of a complementary strand that either had to self-replicate or somehow get replicated by the parental strand or by the duplex, a process with extremely low probability. As emphasized by Christian deDuve, self-replication of RNA "does not make chemical sense" (123) and, indeed, the recent studies do not support this model. RNA molecules can efficiently perform many catalytic functions, such as peptide bond formation, but they make for very poor RNA polymerases [reviewed in (101;102;122;123)]. Alternatively, the first RNA molecules evolved as partners of other molecules, such as peptides, in a variety of early metabolic processes associated with the fusing CCs.

Among the many nucleotide-like compounds (101) that were probably produced in conjunction with early metabolic reactions within the CCs, the ribo-nucleotides had two very important properties that would be critical for the eventual evolution of RNA molecules as catalysts and as informational molecules:

(i) the potential property to form phosphodiester bonds leading to polymeric chains, and

(ii) the property of the nucleotides within the same or different RNA chains to interact with each in a complementary mode facilitated by hydrogen bonds between their bases.

Eventually, the property of the nucleotides to be polymerized into chains and generate linear information was used for generating genetic information, and their property of interacting in a complementary mode was used for replicating this information and for its transfer into three-dimensional information - the proteins. However, likely, the first RNA chains were synthesized for a different purpose that, nonetheless, required the same two critical properties [for a eloquent discussion of this exaptation phenomena in context of molecular evolution see (117)]. In this model, the ribo-nucleotides served as cofactors (a function they still often perform) for diverse catalytic molecules, including (non-specific) peptides assembled in

metabolic complexes within the CCs, facilitating complementary interactions among different metabolic complexes (MC). The formation of phosphodiester bonds among these ribo-nucleotides would generate more stable structural anchors (RNA chains) for the assembly of catalysts in MCs and, more importantly, the formation of complementary hydrogen bonds among RNA chains associated with different MCs would sustain their dynamic, increasingly specific, interactions (Fig.2). These interactions and the selective forces associated with their optimization opened the door for the eventual evolution of MCs into template-based RNA synthesizing machineries. As RNA molecules acquired new structural and catalytic functions, some of these MCs evolved into powerful template-based RNA synthesizing machineries - the ancestral ribosome. The key novelty in this model is that the template-based synthesis of RNA evolved as a mechanism for generating complementary strands (backbones) for the interactive assembly of MCs in a process controlled by natural selection. This view is a full departure from the current evolutionary paradigm that the RNA replication evolved as a "selfish RNA" function, a paradigm that is yet to be reasonably explained (120;123).

The origin and evolution of translation and cellular genome

After the ancestral ribosome was able to reproduce within the CCs by replicating its RNA molecules, it was on the fast track of evolution. The rational for the following model on the evolution of the ancestral ribosome into a specific protein synthesizing machinery (i.e. the origin of translation) is based on the current RNA functions in translation:

(i) Peptide bond formation catalyzed by the RNA moiety (ribozyme) of the large ribosomal subunit;

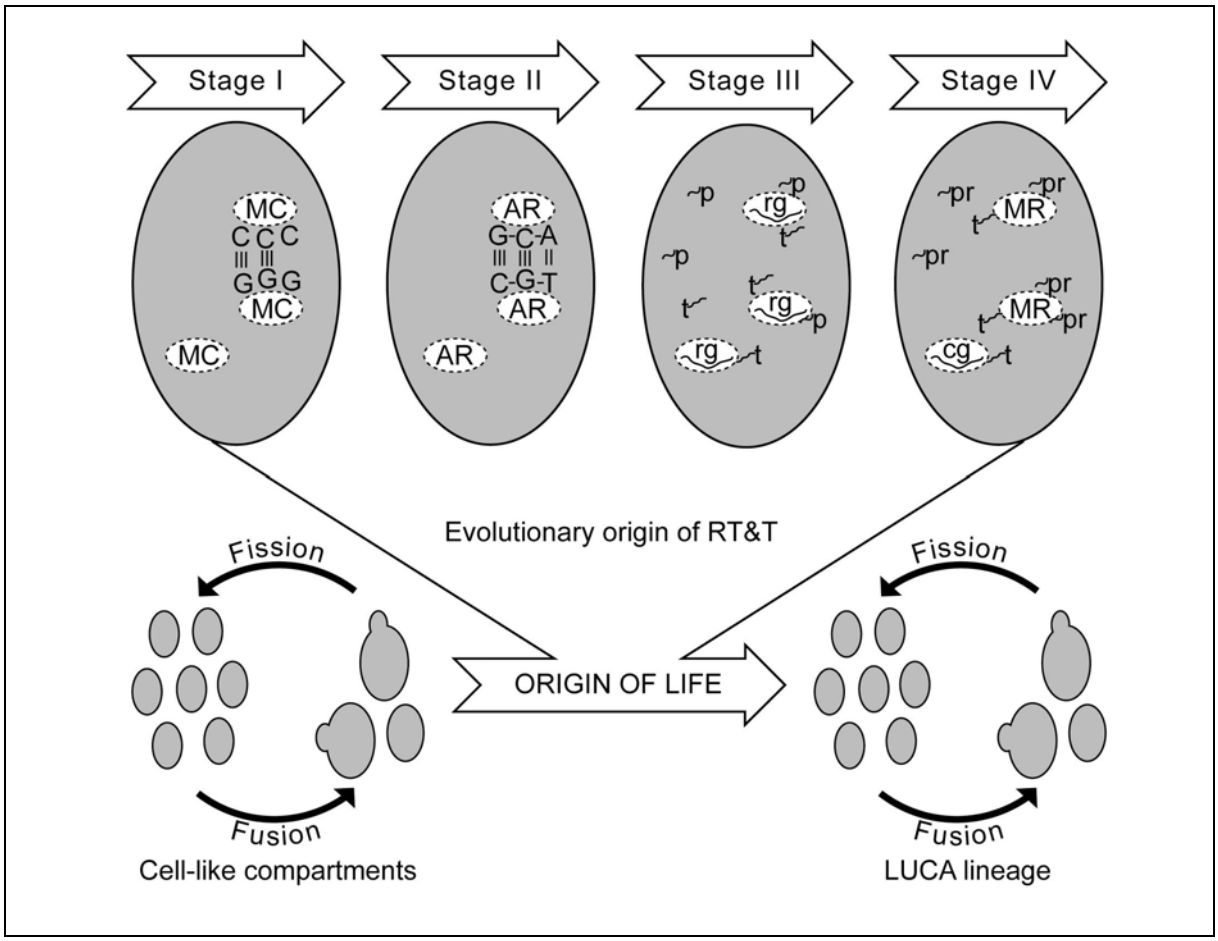

Figure 2. Origin of life. Evolutionary stages in the origin of coupled replication, transcription, and translation system (RT&T) within cell-like compartments that led to their evolutionary transition into the LUCA lineage. Stage I. Evolution of metabolic complexes (MC) that interacted with each other based on complementary hydrogen bonds between ribonucleotides (A, T, C, and G). Stage II. Origin of ancestral ribosomes (AR) as a template-based RNA synthesizing machinery. Stage III. Evolution of ribosomal genomes (rg) as template for the synthesis of various ribosomal RNA molecules [i.e.transcripts (t)]; some of these RNAs catalyzed the synthesis of non-specific peptides (p). Stage VI. Evolutionary transition from ribosomal genomes to a cellular genome (cg) as a centralized unit for storage of genetic information within the cell-like compartments, and the

evolution of ancestral ribosome into a modern ribosome (MR) that synthesized specific proteins (pr) based on genetic code.

(ii) Capture of amino acids by the transfer RNAs (tRNA) involving aminoacylation catalyzed by synthetases; and,

(iii) Presentation of amino acid-charged tRNAs to the peptidyl transferase activity of ribosome using the genetic code, which is based on a three-nucleotide complementary interaction frame between the amino-acid charged tRNAs and a messenger RNA (mRNA).

Among the first efficient enzymatic activities performed by presumed ancestral enzymatic ribosomal RNA molecules (ribozymes) was probably the peptide bond. This RNA enzymatic function is relatively easily selected in laboratory. Although some amino acids were probably incorporated preferentially based on their abundance within the CCs, or based on their "steriochemical fit" within the growing peptide chain, initially, the incorporation order of the amino acids within the ribosomal peptides was more or less random. These early peptides functioned as anchors for the assembly of ribosomal RNA molecules and as cofactors in their enzymatic activity.

Other ribosomal RNA species that evolved relatively early were tRNA-like molecules that steriochemically captured the scarce amino-acids within the fusing CCs and presented them to ribozymes for peptide bond formation. For an optimal presentation, the amino acid-charged tRNA-like molecules interacted with other ribosomal RNA molecules (i.e. ancestral pre-mRNAs) based on multiple, non-contiguous complementary nucleotide interactions. This was the first use of the RNA-based linear information for directing (non-specific) incorporation of amino acids in peptides.

Critical to the evolution of the ancestral ribosome was the process of optimizing the synthesis of ribosomal RNAs and peptides driven by natural selection and sustained by the fusion/fission-based life cycle of the CCs. Very early in the evolution of the ancestral ribosome, the process of replicating individual ribosomal RNA molecules was replaced by a more efficient process involving a ribosomal genome, which was used as a template for the synthesis of diverse ribosomal RNAs. During this early evolutionary stage, each ribosome within the CCs had its own genome for storing and replication of the linear information. Eventually, the ribosomal genome evolved into a "cellular genome," which was an evolutionary breakthrough because it allowed the centralization of linear information storage and replication within the CCs. Following this centralization event, all the ribosomes within a compartment used a single genome for the production their RNA molecules. Free from the burden of storing and replicating RNA, some of the ancestral ribosomal species evolved into powerful specific protein synthesis machineries - the modern ribosome.

Before the evolution of the ancestral "cellular genome" within the CCs, the ancestral ribosomes were already using the linear information encoded by some of the ribosomal RNA molecules for non-specific synthesis of peptides. During the presentation of the amino acids for peptide bond formation, the amino acid-charged tRNA-like molecules interacted with some of the ribosomal RNA molecules (pre-mRNA molecules) based on complementary hydrogen bonds. These complementary interactions opened the door for the evolution of the genetic code. As new species of tRNAs with preferred steriochemical affinity for specific amino acids evolved, their interaction with the ribosomal mRNA-like molecules evolved under the selective pressure of becoming more specific and using less linear information. The optimization of these interactions led to a reduction in the size of the pre-mRNAs, and ultimately to a smaller genome.

To accommodate for 20 amino-acids, the optimal complementary interaction between the charged "transfer RNA" and ribosomal mRNAs was reduced to a reading frame of three contiguous nucleotides. Obviously, if more than 64 different amino acids would have been available for protein synthesis, the minimum reading frame would have been optimized to groups of four nucleotides. Considering the relatively high rate of errors associated with the replication of the ancestral genomes, the specificity among tRNAs/anti-codons and the amino acids evolved to minimize the effect of mutations and translation errors. It is likely, however, that the initial, overall specificity and evolution of the genetic code was established based on the relative abundance of the early amino acids and on the structural and reactive properties of the early peptides within the CCs.

As the ribosomes evolved and were able to synthesize peptides with more diverse and specific order of amino acids, these peptides acquired new functions, both structural and enzymatic, replacing many of ribosomal RNAs. Among the first enzymatic functions performed by the evolving proteins were probably those involved in the replication/transcription of RNA. The rational for this is that, even if RNA molecules would have been able to generate highly diverse three-dimensional structures and enzymatic domains (which based on the chemical properties of RNA is a relatively limited prospect), the potential nucleotide complementary interactions among the presumed RNA templates, the polymerizing RNA-based enzymes, and the RNA products would have precluded the evolution of an efficient RT&T system based primarily on interactions among RNA molecules. The same rational can be used in predicting the extent to which the RNA molecules were presumably involved in other enzymatic activities, such as the myriad of metabolic reactions within the CCs. Therefore, it is likely that the potential intrinsic complementary interactions among a population of diverse RNA molecules precluded the evolution of a true "RNA world" (111).

A similar argument could be used in predicting that the transition from an RNA- to a DNA-based genome probably occurred very early in the evolution of the RT&T. This transition required the evolution of metabolic pathways for the synthesis of dioxy-ribonucleotides, as well as the evolution of an array of other enzymes, including a reverse transcriptase, a DNA polymerase, and a DNA-based RNA polymerase. Although, these specialized polymerases probably evolved from an ancestral RNA template-based polymerase, the selective forces behind their evolution and the mechanism for this transition are elusive. Considering that the RNA has potential reactive properties, a large RNA-based genome might inadvertently develop unstable reactive domains. It is important to realize also that the more complex a RNA genome and associated RT&T, the more difficult it would be to envision selective forces and mechanisms for a transition to a DNA genome.

Therefore, it is highly plausible that a DNA-based genome evolved very early, probably before or during the transition from the "ribosomal genome" to a "cellular genome." Alternatively, this transition occurred later after the hypothetical evolution of true "RNA cells" as described in the model proposed by Patrick Forterre in which the DNA genome and the accompanying enzymes originated in a viral world and later were transferred to RNA cells (95). The critical question in this interesting scenario is what kind of entities were "DNA viruses" and how did they originate in the first place?

Based on the present model for the evolution of coupled RT&T (Fig.2), it is enticing to hypothesize that some ribosomal species, or ribosomal subunits, evolved as virus-like parasitic elements within the fusing CCs. In fact, the ribosomes themselves could be envisioned as reproducing elements (i.e. basically, as molecular organisms) that co-evolved with their hosts - the CCs - and fiercely competed with each other under the laws of natural selection. However, due to the population mode of evolution of CCc driven by their fusion/fission life cycle, the ancestral ribosome co-evolved in a population mode leading to the evolution of the RT&T. And, due to this mode of evolution, it is highly unlikely that some of the ribosomal elements were able to evolve as parasitic species that eventually produced specific transmissible forms such as infectious viral particles (10;11). Also, it is difficult to envision how during this early evolutionary period some of the fusing CCs could have successfully evolved into fusing viral species by becoming reproductively isolated [see the fusion model for the origin of ancestral viruses in (4)]. Reproductive isolation would require specialized receptors in both the host and the parasite, as well as specialized life cycles, which very likely did not evolve before the CCs evolved into true cellular organisms, the LUCA lineage. Probably, only then, some members of this lineage had the opportunity to evolve as parasitic cellular species that fused with their host cells and developed as molecular organisms, the first true ancestral viruses (4).

In conclusion, although highly complex and speculative, this series of models and hypothetical scenarios on the origin and evolution of RT&T system within hypothetical fusing CCs might constitute a plausible integrated conceptual platform for modeling and experimentation leading to laboratory simulation of the critical events in the origin of life.

A new Tree of Life

The evolutionary relationships among the three cellular domains - Archaea, Bacteria and Eukarya - represent the backbone of the current TOL (14-16). Based on evolutionary relationships among ribosomal RNA and other conserved genes, it appears that Archaea and Eukarya are more related to each other; however, the root of the TOL is highly disputed [reviewed in (15;19-27;125)]. Interestingly, the traditional view that the first cellular organisms were Bacteria from which the Archaea and Eukarya eventually evolved, and the challenging view that archaeal/eukaryal lineage originated first are both based on the same sequence data, and on similar phylogenetic analysis programs. Therefore, it appears that the sequence-based phylogenetic approaches might not be able to resolve deep evolutionary relationships going back billions of years without integration into broader evolutionary frameworks.

The goal of TOL is to represent the line of descent among different groups of organisms and species, not necessarily the origin and evolution of their genes [see ref. (4) in which I present multiple TOLs conforming to the major hypotheses on the origin and evolution of viruses]. Due to LGT and differential evolutionary rates of genes, the line of descent cannot always be established based solely on sequence analysis. Particularly, the role played by LGT in cellular evolution is highly disputed (80;126-129).

The viral elements have been one of the main vehicles for LGT and, perhaps evaluating their role in this process might clarify the impact of LGT. This role might also explain the mysterious canonical pattern in the evolution of highly conserved genes, such as ribosomal genes, across the three cellular domains (14;19;80). Despite a relatively long evolutionary period since the three cellular domains originated, the divergence of these genes among the diverse lineages within each domain is lower than that between the three domains, which diverged more recently. This canonical pattern is a highly subtle and unsolved evolutionary issue in cellular evolution (80;95). Based on the co-evolutionary model between viruses and cells presented here, it is very likely the canonical pattern phenomenon is the result of the domain specificity of viral infections and domain-specific LGT process facilitated by these infections. Similar to the population mode of evolution in sexually reproducing species, in which the distribution of alleles is homogenized by sexual reproduction, the intra-domain distribution of genes has been mainstreamed by virus-driven LGT. Because of the paucity of viral infections crossing the domain barrier, the distribution of genes between domains was limited, which has led to the canonical pattern described above. This

model, which is open to modeling experimentation, might be another strong example of the fundamental role played by viruses in cellular evolution.

Despite their evident role in cellular evolution and despite the fact that they are the most abundant life forms, viruses are not included in the TOL. Besides the historical and conceptual reasons that were outlined in a earlier section, as well as in a related paper (4), there seems to be no scientific reason for this exclusion. However, including viruses to the TOL in a conventional way is a difficult task (4;130). First of all, due to their reductive evolution, most viruses have very few genes and, clearly, due to their molecular structure viruses were prone to intense LGT. Considering also the higher number of viral replication errors, larger population size, and short generation time, the viral genes evolved at a much higher rate than cellular genes (32;131;132). For example, at an average rate of nucleotide substitution for an RNA virus of 10^{-4} substitutions per site, per generation, the sequence homology among RNA viral lineages descending from a common ancestor will vanish in a few thousand, or at the most in a few tents of thousands of years (32;131;132). And, although the evolutionally rate of some "DNA viruses" is several orders of magnitude lower, their evolutionary relationship would be difficult to recognize after a relatively short period on the geological time scale. Therefore, the interpretation that the current sequence homology found among diverse extant viral lineages represents evolutionary relationships going back hundreds of million of years or even billions of years (11;133-135), an interpretation that is imposed by the current prevalent view that viruses originated before their cellular hosts, is questionable [see (4;130)].

The new view about the origin and evolution of viruses as molecular organisms (4) dramatically influences the interpretation of current sequenced-based viral phylogenetic studies and offers solutions to many of the issues and questions that arise by trying to explain the current data within the framework of the prevalent current view that viruses originated before cellular lineages. For example, the strong homology found between the DNA polymerase genes of poxviruses and their host cells is currently interpreted as evidence that eukaryal DNA replication machinery and the eukaryal nucleus originated from a poxvirus-like organism (136-139). However, these results are fully compatible with an origin of poxviruses from a parasitic eukaryal species as suggested in the fusion model (4). Another relevant example is in the interpretation of the finding that, despite lack of sequence homology, the viral proteins involved in packaging the viral genome in viruses of different complexity and infecting different cellular domains share functional and structural features (133-135). A popular interpretation of this finding is that these structural and functional similarities are due to a common viral ancestor that predates the origin of cellular domains (133-135), rather than to convergent structural evolution among different viral lineages, or to LGT.

In the unifying evolutionary scenario for the origin and evolution of the viral and cellular domains presented here, the first viral species evolved from the LUCA lineage before the origin of the cellular domains (Fig.1). Although these early viral lineages co-evolved with the ancestral bacterial, archaeal, or eukaryal hosts, it is likely that due to their reductive evolution many viral lineages evolved into symbiotic viruses or went extinct, a likely evolutionary fate of many parasitic viral lineages. Even if some of these early viral lineages did survive evolutionarily as parasites, it is likely that their genes have been replaced through LGT and, therefore, their current genes do not represent their true line of descent, which questions rooting viruses to TOL based strictly on the results of sequence-based phylogenetic analysis [discussed in (4)]. Also, it would be difficult to establish the evolutionarily relationships and the position on the TOL of thousands of new viral lineages that, as predicted by the fusion model, originated from bacterial, archaeal, and eukaryal parasites by reductive evolution throughout the history of life. The broad evolutionary scenario presented here for the evolution of cellular and viral domains supports a TOL in which the LUCA lineage, which originated from cell-like entities as proposed in the "cell-like world" model for the origin of life, is the trunk from which cellular domains and many viral lineages branched off (Fig.1). The order in which the cellular domains originated from the LUCA lineage was probably Bacteria, Archaea, and modern Eukarya. However, many eukaryal features originated probably before the endosymbiotic origin of mitochondria and other plastids, which is commonly considered as the reference evolutionary transition towards modern Eukarya. The topology of this TOL is similar to that put forward by Kandler from a different evolutionary perspective (84;85). The significance and the phylogenetic implications of a phylogenetic tree with this particular topology has been eloquently discussed by Wachtershauser (22;86).

In conclusion, although much of the viral world is yet to be investigated, it is likely that only a few of the extant viral lineages are evolutionary related. It is clear also that many of these viral species cannot be rooted to the TOL in a conventional way. In the present model on the origin and evolution of cellular and viral lineages, most of the TOL are embedded in a viral shell (Fig.1).

Perspective

Outlining the history of life requires broad ideas and scenarios. Metaphorically speaking, only by developing broad evolutionary scenarios, it is possible to see the forest for the trees. And, ultimately, the objective of this article is to draw a fundamental map of the forest generated by life on Earth - a forest that we ironically call the Tree of Life. Certainly, this rough map has gaps and ill-defined boundaries and, obviously, addressing such elusive subjects as the origin and evolution of cellular and viral domains requires speculation; there is no alternative at this time.

In the unifying scenario on the origin and evolution of

cellular and viral domains presented here, viruses drove the evolution of cells, but the cells sustained their own evolution as well as that of the viruses. Before the evolution of the first true cells, and the reductive evolution of some of these cellular lineages into viruses, the universal evolution of the coupled RT&T occurred in self-sustainable CCs. Fundamentally, the rich co-evolutionary history between the cellular and viral domains represents the core of biology and that of the TOL.

Acknowledgments:

I thank Francisco Ayala, William Bains, Eric Bapteste, Rebecca Bandea, Philip Bell, Carolyn Black, Gustavo Caetano-Anollés, Michel Claverie, Patrick Forterre, Nicolas Glansdorff, Johann Gogarten, Kathi Kellar, Eugene Koonin, Norman Pace, Anthony Poole, and Bret Steiner for reviewing this or previous versions of this paper. I also thank Jennifer Hulsey for help with the illustrations.

The findings and conclusions in this report are those of the author and do not necessarily represent the views or the opinions of the U.S. Department of Health and Human Services and Centers for Disease Control and Prevention.

References

1. Hughes,S.S. 1977. *The virus. A History of the Concept.* Heinermann Educational. London.
2. Waterson,A.P., and L.Wilkinson. 1978. *An Introduction to the History of Virology*. Cambridge University Press. London.
3. Knipe,D.M., P.M.Howley, D.E.Griffin, R.A.Lamb, M.A.Martin, B.Roizman, and S.Straus 2007. *Fields Virology*. Lippincott Williams & Wilkins.
4. Bandea,C.I. 2009. The origin and evolution of viruses as molecular organisms. *Nature Precedings* **http://precedings.nature.com/**.
5. Edwards,R.A., and Rohwer,F. 2005. Viral metagenomics. *Nat. Rev. Microbiol.* **3**:504-510.
6. Angly,F.E., Felts,B., Breitbart,M., Salamon,P., Edwards,R.A., Carlson,C., Chan,A.M., Haynes,M., Kelley,S., Liu,H. et al 2006. The marine viromes of four oceanic regions. *PLoS. Biol* **4**:e368.
7. Suttle,C.A. 2007. Marine viruses--major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**:801-812.
8. Claverie,J.M. 2006. Viruses take center stage in cellular evolution. *Genome Biol.* **7**:110.
9. Filee,J., Forterre,P., and Laurent,J. 2003. The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies. *Res. Microbiol.* **154**:237-243.
10. Koonin,E.V., Senkevich,T.G., and Dolja,V.V. 2006. The ancient Virus World and evolution of cells. *Biol. Direct.* **1**:29.
11. Forterre,P. 2006. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* **117**:5-16.
12. Ryan,F.P. 2007. Viruses as symbionts. *Symbiosis* **44**:11-12.
13. Villarreal,L.P. 2005. *Viruses and the Evolution of Life*. ASM Press. Washington.
14. Woese,C.R. 2000. Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci. U. S. A* **97**:8392-8396.
15. Doolittle,W.F. 1999. Phylogenetic classification and the universal tree. *Science* **284**:2124-2129.
16. Lake,J.A., and Rivera,M.C. 2004. Deriving the genomic tree of life in the presence of horizontal gene transfer: conditioned reconstruction. *Mol. Biol Evol.* **21**:681-690.
17. Bandea,C.I. 1983. A new theory on the origin and the nature of viruses. *J. Theor. Biol.* **105**:591-602.
18. Raoult,D., and Forterre,P. 2008. Redefining viruses: lessons from Mimivirus. *Nat. Rev. Microbiol.* **6**:315-319.
19. Woese,C.R. 2002. On the evolution of cells. *Proc. Natl. Acad. Sci. U. S. A* **99**:8742-8747.
20. Martin,W., and Russell,M.J. 2003. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos. Trans. R. Soc. Lond B Biol. Sci.* **358**:59-83.
21. Cavalier-Smith,T. 2006. Cell evolution and Earth history: stasis and revolution. *Philos. Trans. R. Soc. Lond B Biol. Sci.* **361**:969-1006.
22. Wachtershauser,G. 2006. From volcanic origins of chemoautotrophic life to Bacteria, Archaea and Eukarya. *Philos Trans. R. Soc. Lond B Biol Sci* **361**:1787-1806.
23. Penny,D., and Poole,A. 1999. The nature of the last universal common ancestor. *Curr. Opin. Genet. Dev.* **9**:672-677.
24. Doolittle,W.F. 2000. The nature of the universal ancestor and the evolution of the proteome. *Curr. Opin. Struct. Biol* **10**:355-358.
25. Kurland,C.G., Collins,L.J., and Penny,D. 2006. Genomics and the irreducible nature of eukaryote cells. *Science* **312**:1011-1014.
26. Wang,M., Yafremava,L.S., Caetano-Anolles,D., Mittenthal,J.E., and Caetano-Anolles,G. 2007. Reductive evolution of architectural repertoires in proteomes and the birth of the tripartite world. *Genome Res.* **17**:1572-1585.
27. Glansdorff,N., Xu,Y., and Labedan,B. 2008. The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biol Direct.* **3**:29.
28. Forterre,P., and S.Gribaldo 2007. The origin of modern terrestrial life. *HFSP J.* **1**:156-168.
29. Bandea,C.I. 2009. Endogenous viral etiology of prion diseases. *Nature Precedings* **http://precedings.nature.com/**.
30. Morse,S.S. 1994. *The Evolutionary Biology of Viruses*. Raven Press. New York.
31. Gibbs,A.J., Calisher,C.H., and García-Arenal,F. 1995.

Molecular Basis of Virus Evolution. Cambridge University Press. Cambridge.

32. Domingo,E., C.Parrish, and J.J.Holland 2008. *Origin and Evolution of Viruses, 2nd edition*. Elsevier. Oxford.
33. Ryan,F.P. 2009. *Virolution*. Collins. London.
34. Ryan,F.P. 2009. An alternative approach to medical genetics based on modern evolutionary biology. Part 2: retroviral symbiosis. *J R. Soc. Med.* **102**:324-331.
35. Cohan,F.M. 2001. Bacterial species and speciation. *Syst. Biol.* **50**:513-524.
36. Konstantinidis,K.T., and Tiedje,J.M. 2005. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U. S. A* **102**:2567-2572.
37. Fraser,C., Hanage,W.P., and Spratt,B.G. 2007. Recombination and the nature of bacterial speciation. *Science* **315**:476-480.
38. Apanius,V., Penn,D., Slev,P.R., Ruff,L.R., and Potts,W.K. 1997. The nature of selection on the major histocompatibility complex. *Crit Rev Immunol.* **17**:179-224.
39. Hughes,A.L., and Yeager,M. 1998. Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev Genet.* **32**:415-435.
40. Doherty,P.C., and Zinkernagel,R.M. 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* **256**:50-52.
41. Moretta,A. 1997. Molecular mechanisms in cell-mediated cytotoxicity. *Cell* **90**:13-18.
42. Shresta,S., Pham,C.T., Thomas,D.A., Graubert,T.A., and Ley,T.J. 1998. How do cytotoxic lymphocytes kill their targets? *Curr. Opin. Immunol.* **10**:581-587.
43. Freed,E.O., and Martin,M.A. 1995. The role of human immunodeficiency virus type 1 envelope glycoproteins in virus infection. *J Biol Chem.* **270**:23883-23886.
44. Cantin,R., Fortin,J.F., and Tremblay,M. 1996. The amount of host HLA-DR proteins acquired by HIV-1 is virus strain- and cell type-specific. *Virology* **218**:372-381.
45. Kolegraff,K., Bostik,P., and Ansari,A.A. 2006. Characterization and role of lentivirus-associated host proteins. *Exp. Biol Med. (Maywood.)* **231**:252-263.
46. Stott,E.J. 1991. Anti-cell antibody in macaques. *Nature* **353**:393.
47. Arthur,L.O., Bess,J.W., Jr., Sowder,R.C., Benveniste,R.E., Mann,D.L., Chermann,J.C., and Henderson,L.E. 1992. Cellular proteins bound to immunodeficiency viruses: implications for pathogenesis and vaccines. *Science* **258**:1935-1938.
48. Zuckerkandl,E. 2002. Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine. *Genetica* **115**:105-129.
49. Roy,S.W., and Gilbert,W. 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat. Rev. Genet.* **7**:211-221.
50. Rodriguez-Trelles,F., Tarrio,R., and Ayala,F.J. 2006. Origins and evolution of spliceosomal introns. *Annu. Rev. Genet.* **40**:47-76.
51. Lynch,M., and Richardson,A.O. 2002. The evolution of spliceosomal introns. *Curr. Opin. Genet. Dev.* **12**:701-710.
52. Koonin,E.V. 2006. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol Direct.* **1**:22.
53. Jeffares,D.C., Mourier,T., and Penny,D. 2006. The biology of intron gain and loss. *Trends Genet.* **22**:16-22.
54. Taft,R.J., Pheasant,M., and Mattick,J.S. 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**:288-299.
55. Cavalier-Smith,T. 1985. *The Evolution of Genome Size*. Wiley. Chichester.
56. Gregory,T.R. 2004. Insertion-deletion biases and the evolution of genome size. *Gene* **324**:15-34.
57. Brookfield,J.F. 2005. The ecology of the genome - mobile DNA elements and their hosts. *Nat. Rev. Genet.* **6**:128-136.
58. Gogarten,J.P., and Hilario,E. 2006. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC. Evol. Biol* **6**:94.
59. Pennisi,E. 2001. The human genome. *Science* **291**:1177-1180.
60. Beaton,M.J., and Cavalier-Smitht,T. 1999. Eukaryotic non-coding DNA is functional: evidence from the differential scaling of cryptomonad genomes. *Proc. Biol. Sci.* **266**:2053-2059.
61. Hurst,L.D. 1995. Selfish genetic elements and their role in evolution: the evolution of sex and some of what that entails. *Philos Trans. R. Soc. Lond B Biol Sci* **349**:321-332.
62. Orgel,L.E., and Crick,F.H. 1980. Selfish DNA: the ultimate parasite. *Nature* **284**:604-607.
63. Doolittle,W.F., and Sapienza,C. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**:601-603.
64. Bandea,C.I. 1990. A protective function for noncoding, or secondary DNA. *Med. Hypotheses* **31**:33-34.
65. Pozzoli,U., Menozzi,G., Comi,G.P., Cagliani,R., Bresolin,N., and Sironi,M. 2007. Intron size in mammals: complexity comes to terms with economy. *Trends Genet.* **23**:20-24.
66. Mitchell,R.S., Beitzel,B.F., Schroder,A.R., Shinn,P., Chen,H., Berry,C.C., Ecker,J.R., and Bushman,F.D. 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS. Biol.* **2**:E234.
67. Laufs,S., Guenechea,G., Gonzalez-Murillo,A., Zsuzsanna,N.K., Luz,L.M., del,V.C., Jonnakuty,S., Hotz-Wagenblatt,A., Jens,Z.W., Bueren,J.A. et al 2006. Lentiviral vector integration sites in human NOD/SCID

repopulating cells. *J. Gene Med.* **8**:1197-1207.

68. Cattoglio,C., Facchini,G., Sartori,D., Antonelli,A., Miccio,A., Cassani,B., Schmidt,M., von,K.C., Howe,S., Thrasher,A.J. et al 2007. Hot spots of retroviral integration in human CD34+ hematopoietic cells. *Blood* **110**:1770-1778.
69. Cavalier-Smith,T. 2005. Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. *Ann. Bot. (Lond)* **95**:147-175.
70. Lewinski,M.K., and Bushman,F.D. 2005. Retroviral DNA integration--mechanism and consequences. *Adv. Genet.* **55**:147-181.
71. Uren,A.G., Kool,J., Berns,A., and van Lohuizen M. 2005. Retroviral insertional mutagenesis: past, present and future. *Oncogene* **24**:7656-7672.
72. Bushman,F.D. 2007. Retroviral integration and human gene therapy. *J. Clin. Invest* **117**:2083-2086.
73. Amariglio,N., and Rechavi,G. 1993. Insertional mutagenesis by transposable elements in the mammalian genome. *Environ. Mol. Mutagen.* **21**:212-218.
74. Craig,N.L. 1997. Target site selection in transposition. *Annu. Rev. Biochem.* **66**:437-474.
75. Makarova,K.S., Grishin,N.V., Shabalina,S.A., Wolf,Y.I., and Koonin,E.V. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct.* **1**:7.
76. Bagasra,O., and Prilliman,K.R. 2004. RNA interference: the molecular immune system. *J. Mol. Histol.* **35**:545-553.
77. Sorek,R., Kunin,V., and Hugenholtz,P. 2008. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* **6**:181-186.
78. Cerutti,H., and Casas-Mollano,J.A. 2006. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr. Genet.* **50**:81-99.
79. Cavalier-Smith,T. 2002. Origins of the machinery of recombination and sex. *Heredity* **88**:125-141.
80. Woese,C.R. 2004. A new biology for a new century. *Microbiol. Mol. Biol. Rev.* **68**:173-186.
81. Woese,C. 1998. The universal ancestor. *Proc. Natl. Acad. Sci. U. S. A* **95**:6854-6859.
82. de Duve C. 2005. The onset of selection. *Nature* **433**:581-582.
83. Ayala,F.J. 1998. Is sex better? Parasites say "no". *Proc. Natl. Acad. Sci. U. S. A* **95**:3346-3348.
84. Kandler,O. 1994. Cell wall biochemistry in Archaea and its phylogenetic implications. *J Biol Phys* **20**:165-169.
85. Kandler,O. 1994. The early diversification of life. In *Nobel Symposium No. 84. Early Life on Earth*. S.Bengtson, editor. Columbia University Press. New York. 152-160.
86. Wachtershauser,G. 2003. From pre-cells to Eukarya--a tale of two lipids. *Mol. Microbiol.* **47**:13-22.
87. Kandler,O., and Konig,H. 1998. Cell wall polymers in Archaea (Archaebacteria). *Cell Mol. Life Sci.* **54**:305-308.
88. Pereto,J., Lopez-Garcia,P., and Moreira,D. 2004. Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem. Sci* **29**:469-477.
89. Cavalier-Smith,T. 2001. Obcells as proto-organisms: membrane heredity, lithophosphorylation, and the origins of the genetic code, the first cells, and photosynthesis. *J. Mol. Evol.* **53**:555-595.
90. Wachtershauser,G. 2007. On the chemistry and evolution of the pioneer organism. *Chem. Biodivers.* **4**:584-602.
91. Boucher,Y., Kamekura,M., and Doolittle,W.F. 2004. Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol. Microbiol.* **52**:515-527.
92. de Duve C. 2007. The origin of eukaryotes: a reappraisal. *Nat. Rev. Genet.* **8**:395-403.
93. Poole,A.M., and Penny,D. 2007. Evaluating hypotheses for the origin of eukaryotes. *Bioessays* **29**:74-84.
94. Embley,T.M., and Martin,W. 2006. Eukaryotic evolution, changes and challenges. *Nature* **440**:623-630.
95. Forterre,P. 2006. Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc. Natl. Acad. Sci. U. S. A* **103**:3669-3674.
96. Margulis,L., Chapman,M., and Dolan,M.F. 2007. Semes for analysis of evolution: de Duve's peroxisomes and Meyer's hydrogenases in the sulphurous Proterozoic eon. *Nat. Rev. Genet.* **8**:1.
97. Gray,M.W., Burger,G., and Lang,B.F. 1999. Mitochondrial evolution. *Science* **283**:1476-1481.
98. Martin,W., and Koonin,E.V. 2006. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* **440**:41-45.
99. de Duve C.and S.L.Miller 1991. Two-dimensional life? *Proc. Natl. Acad. Sci. U. S. A* **88**:10014-10017.
100. Wachtershauser,G. 1994. Life in a ligand sphere. *Proc. Natl. Acad. Sci. U. S. A* **91**:4283-4287.
101. Orgel,L.E. 2004. Prebiotic chemistry and the origin of the RNA world. *Crit Rev. Biochem. Mol. Biol.* **39**:99-123.
102. Joyce,G.F. 2002. The antiquity of RNA-based evolution. *Nature* **418**:214-221.
103. Deamer,D.W. 1997. The first living systems: a bioenergetic perspective. *Microbiol. Mol. Biol. Rev.* **61**:239-261.
104. Segre,D., Ben-Eli,D., Deamer,D.W., and Lancet,D. 2001. The lipid world. *Orig. Life Evol. Biosph.* **31**:119-145.
105. Copley,S.D., Smith,E., and Morowitz,H.J. 2007. The origin of the RNA world: co-evolution of genes and metabolism. *Bioorg. Chem.* **35**:430-443.

106. Fry,I. 2006. The origins of research into the origins of life. *Endeavour* **30**:24-28.
107. Shapiro,R. 2006. Small molecule interactions were central to the origin of life. *Q. Rev. Biol* **81**:105-125.
108. Lahav,N., Nir,S., and Elitzur,A.C. 2001. The emergence of life on Earth. *Prog. Biophys. Mol. Biol* **75**:75-120.
109. Hunding,A., Kepes,F., Lancet,D., Minsky,A., Norris,V., Raine,D., Sriram,K., and Root-Bernstein,R. 2006. Compositional complementarity and prebiotic ecology in the origin of life. *Bioessays* **28**:399-412.
110. Koonin,E.V., and Martin,W. 2005. On the origin of genomes and cells within inorganic compartments. *Trends Genet.* **21**:647-654.
111. Gilbert,W. 1986. Origin of life, the RNA world. *Nature* **319**:618.
112. Hanczyc,M.M., and Szostak,J.W. 2004. Replicating vesicles as models of primitive cell growth and division. *Curr. Opin. Chem. Biol* **8**:660-664.
113. Walde,P. 2006. Surfactant assemblies and their various possible roles for the origin(s) of life. *Orig. Life Evol. Biosph.* **36**:109-150.
114. Monnard,P.A., and Deamer,D.W. 2002. Membrane self-assembly processes: steps toward the first cellular life. *Anat. Rec.* **268**:196-207.
115. Hanczyc,M.M., Fujikawa,S.M., and Szostak,J.W. 2003. Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science* **302**:618-622.
116. Hanczyc,M.M., Mansy,S.S., and Szostak,J.W. 2007. Mineral surface directed membrane assembly. *Orig. Life Evol. Biosph.* **37**:67-82.
117. Wolf,Y.I., and Koonin,E.V. 2007. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biol. Direct.* **2**:14.
118. Woese,C.R. 2001. Translation: in retrospect and prospect. *RNA.* **7**:1055-1067.
119. Noller,H.F. 2004. The driving force for molecular evolution of translation. *RNA.* **10**:1833-1837.
120. Koonin,E.V. 2007. The cosmological model of eternal inflation and the transition from chance to biological evolution in the history of life. *Biol Direct.* **2**:15.
121. Poole,A.M., Jeffares,D.C., and Penny,D. 1998. The path from the RNA world. *J. Mol. Evol.* **46**:1-17.
122. Muller,U.F. 2006. Re-creating an RNA world. *Cell Mol. Life Sci* **63**:1278-1293.
123. de Duve C. 2007. Chemistry and selection. *Chem. Biodivers.* **4**:574-583.
124. Pace,N.R. 2001. The universal nature of biochemistry. *Proc. Natl. Acad. Sci. U. S. A* **98**:805-808.
125. Forterre,P., and Philippe,H. 1999. Where is the root of the universal tree of life? *Bioessays* **21**:871-879.
126. Kurland,C.G., Canback,B., and Berg,O.G. 2003. Horizontal gene transfer: a critical view. *Proc. Natl. Acad. Sci. U. S. A* **100**:9658-9662.
127. Gogarten,J.P., and Townsend,J.P. 2005. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* **3**:679-687.
128. Bapteste,E., Boucher,Y., Leigh,J., and Doolittle,W.F. 2004. Phylogenetic reconstruction and lateral gene transfer. *Trends Microbiol.* **12**:406-411.
129. Rivera,M.C., and Lake,J.A. 2004. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431**:152-155.
130. Moreira,D., and Lopez-Garcia,P. 2009. Ten reasons to exclude viruses from the tree of life. *Nat. Rev. Microbiol.* **7**:306-311.
131. Holmes,E.C. 2003. Molecular clocks and the puzzle of RNA virus origins. *J. Virol.* **77**:3893-3897.
132. Desjardins,C., Eisen,J.A., and Nene,V. 2005. New evolutionary frontiers from unusual virus genomes. *Genome Biol* **6**:212.
133. Bamford,D.H., Grimes,J.M., and Stuart,D.I. 2005. What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol* **15**:655-663.
134. Hendrix,R.W., Lawrence,J.G., Hatfull,G.F., and Casjens,S. 2000. The origins and ongoing evolution of viruses. *Trends Microbiol.* **8**:504-508.
135. Koonin,E.V., and Dolja,V.V. 2006. Evolution of complexity in the viral world: the dawn of a new vision. *Virus Res.* **117**:1-4.
136. Villarreal,L.P., and DeFilippis,V.R. 2000. A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J. Virol.* **74**:7079-7084.
137. Takemura,M. 2001. Poxviruses and the origin of the eukaryotic nucleus. *J. Mol. Evol.* **52**:419-425.
138. Bell,P.J. 2001. Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J. Mol. Evol.* **53**:251-256.
139. Bell,P.J. 2006. Sex and the eukaryotic cell cycle is consistent with a viral ancestry for the eukaryotic nucleus. *J. Theor. Biol* **243**:54-63.