ELSEVIER

Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

1

Introduction to virus origins and their role in biological evolution

Abbreviations

AIDS acquired immune deficiency syndrome
APOBEC apolipoprotein B mRNA editing complex
CCMV cowpea chlorotic mottle virus
dsRNA double-stranded RNA
E. coli Escherichia coli
eHBV endogenous hepatitis B viruses
HBV hepatitis B virus
HCV hepatitis C virus
HDV hepatitis delta virus
HIV-1 human immunodeficiency virus type 1
ICTV International Committee on Taxonomy of Viruses
Kbp thousand base pairs
mRNA messenger RNA
PMWS postweaning multisystemic wasting syndrome
RT reverse transcriptase
RdRp RNA-dependent RNA polymerase
ssRNA single-stranded RNA
T7 bacteriophage T7
tRNA transfer RNA
UV ultraviolet

1.1 Considerations on biological diversity

To approach the behavior of viruses acting as populations, we must first examine the diversity of the present-day biosphere and the physical and biological context in which primitive viral forms might have arisen. Evolution pervades nature. Thanks to new theories and to the availability of powerful instruments, new experimental procedures, and increasing computing power—which together constitute the very roots of scientific progress—we know that the physical and biological worlds are constantly evolving. Several classes of energy have gradually shaped matter and living entities, basically as the outcome of random events and Darwinian natural selection in its broadest sense. The identification of DNA as the genetic material and the advent of genomics in the second half of the twentieth century unveiled an astonishing degree of diversity within the living world that derives mainly from combinations of four classes of nucleotides. Biodiversity, a term coined by O. Wilson in 1984 and emphasized by T. Lovejoy and others, is a feature of all living beings, be differentiated multicellular organisms, single-cell organisms, or subcellular genetic elements, among them the viruses. Next-generation sequencing methods developed at the beginning of the 21st century allow thousands of sequences from the same biological sample (a microbial community in a soil or ocean sample, a tumor, or an infected host) to be determined. These procedures have documented the presence of

myriads of variants in a "single biological entity" or in "communities of biological entities." Differences extend to individuals that belong to the same biological group, be it *Homo sapiens*, *Drosophila melanogaster*, *Escherichia coli*, or human immunodeficiency virus type 1 (HIV-1). No exceptions have been described. Diversity is extensive and not restricted to the genotypic level. It also affects phenotypic traits.

During decades, in the first half of the twentieth century, population genetics had as one of its tenets that genetic variation due to mutation had for the most part been originated in a remote past. It was generally thought that the present-day diversity was essentially brought about by the reassortment of chromosomes during sexual reproduction. This view was weakened by the discovery of extensive genetic polymorphisms, first in *Drosophila* and humans, through secondary analyses of electrophoretic mobility of enzymes, detected by in situ activity assays to yield zymograms that were displayed as electromorphs. These early studies on allozymes were soon extended to other organisms. Assuming that no protein modifications had specifically occurred in some individuals, the results suggested the presence of several different (allelic) forms of a given gene among individuals of the same species, be it humans, insects, or bacteria. In the absence of information on DNA nucleotide sequences, the first estimates of heterogeneity from the numbers of electromorphs were collated with the protein sequence information available. An excellent review of these developments (Selander, 1976) ended with the following premonitory sentence on the role of molecular biology in unveiling evolutionarily relevant information: "Considering the magnitude of this effect, we may not be overfanciful to think that future historians will see molecular biology more as the salvation for than, as it first seemed, the nemesis of evolutionary biology."

The conceptual break was confirmed and accentuated when molecular cloning and nucleotide sequencing techniques produced genomic nucleotide sequences from multiple individuals of the same biological species. Variety has shaken our classification schemes, opening a debate on how to define and delimit biological "species" in the microbial world. From a medical perspective, it has opened the way to "personalized" medicine, so different are the individual contexts in which disease processes (infectious or other) unfold. Diversity is a general feature of the biological world, with multiple implications for interactions in the environment, and also for human health and disease (Bernstein, 2014).

1.2 Some questions of current virology and the scope of this book

Viruses (from the Latin "virus," poison) are no exception regarding diversity. The number of different viruses and their dissimilarity in shape and behavior is astounding. Current estimates indicate that the total number of virus particles in our biosphere reaches 10^{32}, exceeding by one order of magnitude the total number of cells. Viruses are found in surface and deep-sea and lake waters, below the Earth surface, in any type of soil, in deserts, and in most environments designated as extreme regarding ionic conditions (i.e., hypersaline) and temperature (thermophilic) (Breitbart et al., 2004; Villarreal, 2005; Lopez-Bueno et al., 2009; Box 1.1). The viruses that have been studied are probably a minimal and biased representation of those that exist, with at least hundred thousand mammalian viruses awaiting discovery (Anthony et al., 2013; Epstein and Anthony, 2017). This is because high-throughput screening procedures have only recently become available, and also because prevention of disease has provided the main incentive to study viruses. Disease-associated viruses are those most described in the scientific literature.

BOX 1.1

SOME NUMBERS CONCERNING VIRUSES IN THE EARTH BIOSPHERE

- Total number of viral particles: $\sim 10^{32}$. This is 10 times more than cells, and they are equivalent to 2×10^8 tons of carbon.
- Virus particles in 1 cm^3 of seawater: $\sim 10^8$.
- Virus particles in 1 m^3 of air: $\sim 2 \times 10^6$ to 40×10^6.
- Rate of viral infections in the oceans: $\sim 1 \times 10^{23}$/s.

- A string with the viruses on Earth would be about $\sim 2 \times 10^8$ light-years long ($\sim 1.9 \times 10^{24}$ m). This is the distance from Earth of the galaxy clusters Centaurus, Hydra, and Virgo.

Based on: Suttle (2007), Whon et al. (2012), and Koonin and Dolja (2013).

Current virology poses some general and fascinating questions, which are not easily approachable experimentally. Here are some:

- What is the origin of viruses?
- Did they originate before or after a cellular world was in place?
- What selective forces have maintained multiple viruses as parasites of unicellular and multicellular organisms?
- Do viruses exist essentially as selfish parasites, or have they played constructive roles in the biosphere?
- Why have a few viral forms not outcompeted most other forms? Or have they?
- Have viruses been maintained as modulators of the population numbers of their host species?
- Does virus variation play a role in the unfolding of viral disease processes?
- Is it quasispecies dynamics that characterize RNA and many DNA viruses, a remnant of an adaptive strategy that presided all life forms in the remote past?

- Is the behavior of present-day viruses at the population level only an inheritance of their origins or a present-day necessity?

This book deals with some of these issues, mainly those that are amenable to experimental testing. Topics covered include molecular mechanisms of genetic variation, with emphasis on high mutation rates, Darwinian principles acting on viruses, quasispecies dynamics and its implications, consequences for virus-host interactions, fitness as a relevant parameter, experimental model systems in cell culture, ex-vivo and in vivo, long-term virus evolution, the current situation of antiviral strategies to confront quasispecies swarms, and conceptual extensions of quasispecies to nonviral systems. These subjects have as a common thread that Darwinian natural selection has an immediate imprint on them, observable in the time scale of days or even hours. The capacity for rapid evolution displayed by viruses represents an unprecedented and often underappreciated development in biology: the direct observation of Darwinian principles at play within short times.

Evolution is defined as a change in the genetic composition of a population over time. In this book evolution will be used in its broader sense to mean any change in the genetic composition of a virus over time, irrespective of the time frame involved, and the transience of the change. We treat as evolution both the genome variation that poliovirus has undergone from the middle of the 20th century to present days and the changes that the same virus undergoes within an infected individual. It is remarkable that only a few decades ago virus evolution (or for that matter microbial evolution in general) was not considered a significant factor in viral pathogenesis. Evolution was largely overlooked in the planning of strategies for microbial disease control. A lucid historical account of the different perceptions of virus evolution, including early evidence of phenotypic variation of viruses, with emphasis on the impact of the complexity of RNA virus populations, was written by J.J. Holland (2006). Despite having been largely ignored by virologists, the present book was partly stimulated by the conviction that the concept of complexity is pertinent to the understanding of viruses at the population level, having direct connections with viral disease and disease control.

1.3 The staggering ubiquity and diversity of viruses: limited morphotypes

Despite pleomorphism in cells and viruses (presence or not of envelopes, and viruses being spherical or even displaying a lemon-like shape, or being elongated), the size of viral particles and their host cells tends to be commensurate with the amount of genetic material that they contain and transmit to progeny (Fig. 1.1). The 2017 report from International Committee on Taxonomy of Viruses (ICTV) divides viruses into 10

Orders, and each of them is subdivided into several families, subfamilies, genera, species, and isolates (https://talk.ictvonline.org/), and each isolate includes a multitude of variants. The task of classifying viruses meets with considerable hurdles and requires periodic revisions by the ICTV, an organization whose role has been essential to provide conceptual order in the vast viral world. One of its objectives is the assignment of newly discovered viruses to the adequate group. A remarkable number of isolates remain unclassified, an echo of the
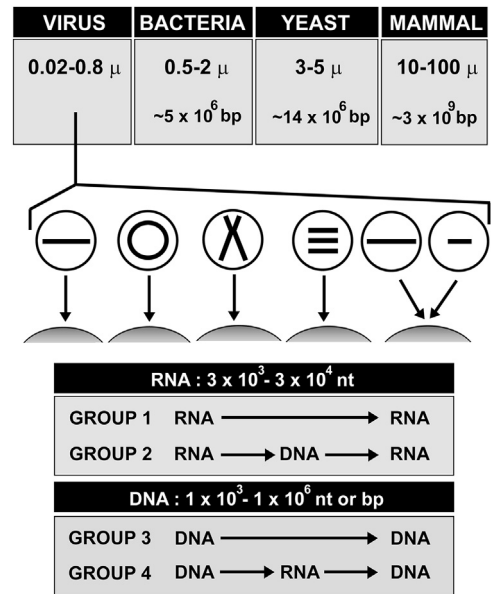

VIRUS	BACTERIA	YEAST	MAMMAL
0.02-0.8 μ	0.5-2 μ	3-5 μ	10-100 μ
	$\sim 5 \times 10^6$ bp	$\sim 14 \times 10^6$ bp	$\sim 3 \times 10^9$ bp

RNA : 3×10^3 - 3×10^4 nt

| GROUP 1 | RNA $\longrightarrow$ RNA |
| GROUP 2 | RNA $\longrightarrow$ DNA $\longrightarrow$ RNA |

DNA : 1×10^3 - 1×10^6 nt or bp

| GROUP 3 | DNA $\longrightarrow$ DNA |
| GROUP 4 | DNA $\longrightarrow$ RNA $\longrightarrow$ DNA |

FIGURE 1.1   Representative average diameter values and genome complexity of viruses and some cell types. Diameters are expressed in microns (μ), length of DNA in base pairs (bp), and of RNA in nucleotides (nt). Viral genomes can be linear, circular, diploid, segmented, or bipartite (multipartite in general; genome segments encapsidated in separate particles); in the latter case at least two particles, each with one kind of genomic segment, must infect the same cell for progeny production. The bottom boxes describe four groups of viruses according to the type of nucleic acid that acts as replicative intermediate.

natural diversity of viruses, even among the limited subset that has been isolated and characterized.

Viruses can be divided into two broad groups: those that have RNA as genetic material, termed the *RNA viruses*, and those that have DNA as genetic material termed the *DNA viruses*. They can have linear, circular, or segmented genomes of single-stranded or double-stranded nucleic acid (Fig. 1.1). All evidence suggests that the polynucleotide chain (or chains) that constitute the viral genome has all the information to generate infectious progeny in a cell, as evidenced by the production of infectious poliovirus from synthetic DNA copies assembled to represent the genomic nucleotide sequence (Cello et al., 2002).

With regard to the concepts of genome stability versus variation addressed in this book, it is helpful to divide viruses into four groups, depending on whether it is DNA or RNA the type of genetic material, which acts as a replicative intermediate in the infected cell (bottom gray shaded boxes in Fig. 1.1). The nucleic acids written in the four schemes are those involved in the flow of genetic information (indicated by arrows), not in gene expression since all of them use messenger RNA (mRNA) for virus-specific protein synthesis. Mistakes in the form of misincorporation of nucleotides during the replication steps indicated by arrows are transmitted to progeny genomes. RNAs produced by transcription to serve solely as mRNAs are essential for gene expression and virus multiplication, but misincorporations in such transcripts are not transmitted to progeny. It could be considered that some mRNA molecules when synthesized from the corresponding RNA or DNA template may acquire mutations and that these mutated molecules (e.g., an mRNA encoding a viral polymerase), when translated, may produce a polymerase with lower copying fidelity that will evoke additional mutations; we will ignore this

possibility since a single mRNA molecule should have a rather limited contribution to the overall genetic variation of a replicating virus population with thousands of polymerase and template molecules in the replication complexes (or replication factories).

Group 1 (with a replicative scheme abbreviated as RNA→RNA) includes RNA viruses whose genomic replication cycle involves only RNA. They are sometimes called *riboviruses*. Examples are the influenza viruses, hepatitis A and C viruses, poliovirus, coronaviruses, Ebola virus, foot-and-mouth disease virus, or tobacco mosaic virus, among many other important human, animal, and plant pathogens. Their replication is catalyzed by an RNA-dependent RNA polymerase (RdRp) encoded in the viral genome, often organized as a replication complex with viral and host proteins in cellular membrane structures.

Group 2 (RNA→DNA→RNA) comprises the retroviruses [such as HIV-1, the acquired immune deficiency syndrome (AIDS) virus, and several tumor viruses] that retrotranscribe their RNA into DNA. Retrotranscription is catalyzed by reverse transcriptase (RT), an RNA-dependent DNA polymerase encoded in the retroviral genome. It reverses the first step in the normal flow of expression of genetic information from DNA to RNA to protein, once known as the *dogma* of molecular biology. This enzyme was instrumental in the understanding of cancer, and for genetic engineering and the origin of modern biotechnology. As a historical account of the impact of H. Temin's work (codiscoverer of RT with D. Baltimore), the reader is referred to Cooper et al. (1995). Retroviruses include a provirus stage in which the viral DNA is integrated into host DNA. When silently installed in cellular DNA, the viral genome behaves mostly as a cellular gene.

Group 3 (DNA→DNA) contains most DNA viruses, such as herpesviruses, poxviruses,

iridoviruses, and papillomaviruses, and the extremely large viruses of amebae (i.e., Mimivirus, Megavirus, or Pandoravirus, generically termed *giant viruses*) (La Scola et al., 2008; Colson et al., 2017). Their replication is catalyzed by a DNA-dependent DNA polymerase either encoded in the viral genome or in the cellular DNA. Cellular DNA polymerases are involved in the replication of DNA viruses that do not encode their own DNA polymerase.

Finally, Group 4 (DNA→RNA→DNA) includes viruses which despite having DNA as genetic material, produce an RNA as a replicative intermediate, the most significant examples being the human and animal hepatitis B viruses (HBVs) and the cauliflower mosaic virus of plants, termed *hepadnaviruses*.

Most viruses, from the more complex DNA viruses [i.e., 1200 Kbp (1000 base pairs) for the ameba Mimivirus, 752 Kbp for some tailed bacteriophages, and up to 370 Kbp for poxviruses, iridoviruses, and herpesviruses], the virophages that are parasites of the giant DNA viruses, the simplest DNA viruses (the circular single-stranded 1760 residue DNA of porcine circovirus), RNA bacteriophages (4220 nucleotides of ssRNA for bacteriophage Qβ), or subviral elements (viroids, virusoids, satellites, and helper-dependent defective replicons) show remarkable genetic diversity. However, RNA viruses that replicate entirely via RNA templates (Group 1 in Fig. 1.1); retroviruses (Group 2); and the hepadnaviruses (Group 4) display salient genetic plasticity, mainly in the way of a high rate of introduction of point mutations (Chapter 2). Their mutability may be an inheritance of universal flexibility that probably characterized primitive RNA or RNA-like molecules, thought to have populated an ancestral RNA world at an early stage of life on Earth (Section 1.4.2). Thus, the presence of RNA at any place in the replicative schemes (Group 1, 2, and 4 in Fig. 1.1) implies error-prone replication and the potential of very rapid evolution. "Potential" must be underlined because high error rates do not necessarily result in rapid long-term evolution in nature (Chapter 7).

The extent of genetic variation and its biological consequences have been less investigated for DNA viruses than for RNA viruses. The available data suggest that DNA viruses are closer to RNA viruses than suspected only a few years ago, regarding their capacity of variation and adaptation. This is particularly true of the single-stranded DNA viruses of animals and plants. Evolutionary theory predicts that high-fidelity polymerase machinery is necessary to maintain the informational stability of complex genomes (those that carry a large amount of genetic information). This necessity accomplished by proofreading-repair and postreplicative-repair activities that assist the replicative DNA polymerases and their cellular and viral DNA progeny.

Our current capacity to sample many thousands viral genomes in short times (a trend that is continuously expanding) is revealing an astonishing number of slightly different viral genomes within a single infected host, and even within an organ or within individual cells of an organ! Intrahost diversity of viruses can be the result not only of diversification within the host but also of coinfection with different viruses (or variants of one virus), or an infection that triggers reactivation of a related or unrelated virus from a latent reservoir, or combined effects of these mechanisms. In turn, interhost (long-term) virus diversification can result from selection acting on variants generated by mutation, recombination or reassortment, and random sampling events (independent of selection) within hosts and in host-to-host transmission, or their combined effects (Chapters 2 and 3).

A different picture of diversity is obtained by comparing the morphological characteristics of viral particles (termed *virions*). The hundreds of

thousands of bacterial and archaeal viruses that have been recognized can be assigned to as few as 20 morphotypes. The capsids of nonenveloped (naked) viruses display helical or icosahedral symmetry that determines the architecture of the virion. Variation in size and surface protein distributions can be attained from limited protein folds and the same symmetry principles (Mateu, 2013, Fig. 1.2). Divergent primary amino acid sequences in proteins can fold in closely related structures. The "structural space"

available to viruses particles is much more restricted than the "sequence space" available to viral genomes (Abrescia et al., 2012). Sequence space and its mapping into a phenotypic space are key concepts for the understanding of evolutionary mechanisms (Chapter 3).

Three-dimensional structures of entire virions or their constituent proteins can provide an overview of phylogenetic lineages and evolutionary steps in cases in which the information cannot be attained by viral genomics (Ravantti et al.,
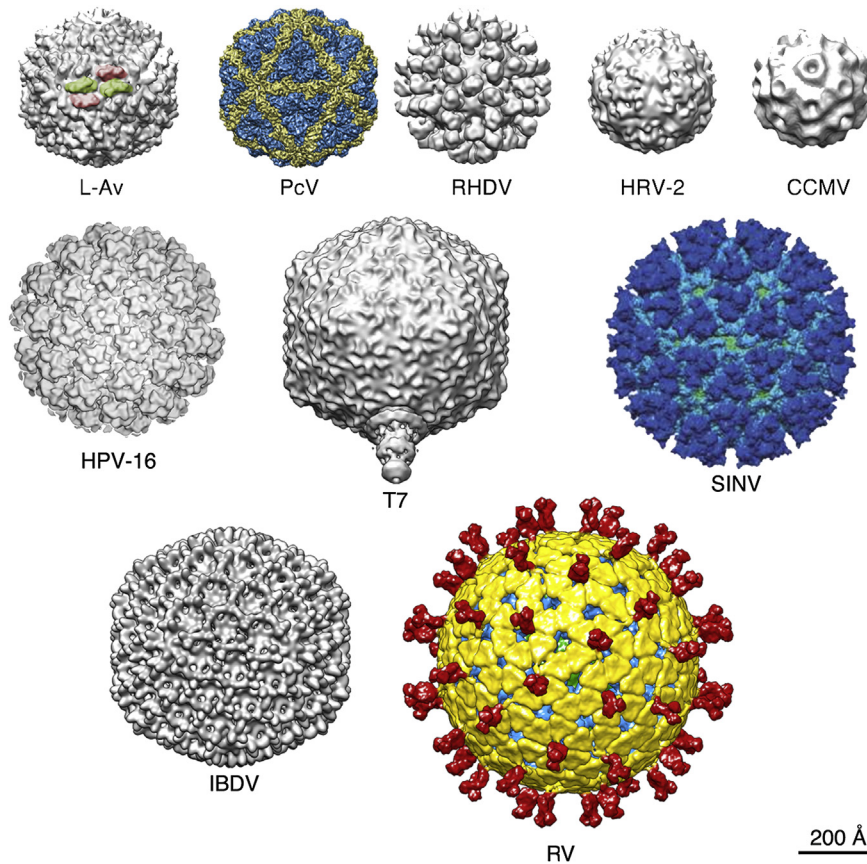
FIGURE 1.2   Examples of spherical bacterial, fungal, plant, and animal virus particles reconstructed from cryo-electron microscopy images. L-Av, *Saccharomyces cerevisae* virus; PcV, *Penicillum Chrysogenum* virus; RHDV, rabbit hemorrhagic disease virus; HRV-2 human rhinovirus type 2; CCMV, cowpea chlorotic mottle virus; HPV-16, human papillomavirus type 16; T7, head of bacteriophage T7; SINV, Sindbis virus; IBVD, Infectious bursal disease virus; RV, rotavirus. *Picture modified from one kindly supplied by J.R. Castón, Carrascosa, J.L., 2013. The basic architecture of viruses. In: Mateu, M.G. (Ed.), Structure and Physics of Viruses. Springer, Dordrecht, Heidelberg, New York, London, pp. 53–75 with permission.*

2013). Yet, minor genetic modifications that do not affect the phylogenetic position of a virus or the structure of the encoded viral proteins in any substantial manner can nevertheless have major consequences for traits as important as host range or pathogenicity. How such minor changes in viruses can have major biological consequences may relate to the historical role of viruses in an evolving, ancestral, pre-cellular biosphere. To further address this issue, we need to examine how viruses may have originated. This, in turn, begs the question of the origin of life and the possible involvement of viruses in early life development.

1.4 Origin of life: a brief historical account and current views

An understanding of the mechanisms involved in the origin of life may help in penetrating into the origin of previral entities, defined as the precursors of the viruses we isolate in modern times. Different notions on the origin of life have been held in human history, often linked to religious debate. Opinions have ranged from a conviction of the spontaneous and easy generation of life from inanimate materials, or its beginning from a unique and rare combination of small prebiotic molecules, or being the result of a lengthy prebiotic process, or its inevitability as the outcome of the evolution of matter in our universe (or "sets" of universes with the adequate physical parameters, according to some cosmological models).

As little as 150 years ago (a time not distant from the discovery of the first viruses), there was a general belief in the spontaneous generation of life. This was somewhat paradoxical because chemists of the seventeenth century divided chemistry into mineral chemistry, vegetal chemistry, and animal chemistry. J.J. Berzelius put together animal and plant chemistry

and named the resulting discipline "organic" chemistry, which he distinguished from "inorganic" chemistry (Berzelius, 1806). He formulated what was known as the *central dogma of chemistry*: "The generation of organic compounds from inorganic compounds, outside a living organism, is impossible." The classical experiments of L. Pasteur provided definitive proof that, at least under the prevailing conditions on present-day Earth, "life comes from life" (Pasteur, 1861). He established what was considered the "central dogma of biology": "The generation of a whole living organism from chemical compounds, outside a living organism is impossible." The requirement of life to generate life was, however, extended to the belief that "living" and "nonliving" were two separate categories in the organization of matter, and that organic compounds could be synthesized only by living cells. This doctrine, called *vitalism* dominated biology for almost a century, and in a modified manner, it continues today regarding the interpretation of mental activity in humans (matter and spirit as "substance dualism"). Historical views on the origin of life have been addressed in several publications (Rohlfing and Oparin, 1972; Bengtson, 1994; de Duve, 2002; Eigen, 2002, 2013; Lazcano, 2010).

Dogmas are generally not to stay. "Vitalism" was shattered by the chemical synthesis of organic compounds from inorganic precursors (urea by F. Wöhler in 1828, acetic acid by H. Kolbe in 1845, hydrocarbons by D. Mendeleev in 1877, and several other compounds by M. Berthelot in the second half of the 19th century). The evidence that no "vital force" was needed for such syntheses led F.A. Kekulé to write in his classic textbook on organic chemistry published during 1859—60: "We have come to the conviction that … no difference exists between organic and inorganic compounds." From then on, organic chemistry became the chemistry of "carbon compounds." We know now that living entities are made of the same chemical elements

found in the mineral world. Of the 118 elements in the periodic table (in 2019), 59 are found in the human body.

A key experiment carried out in 1953 by S. Miller, working with H.C. Urey showed that components of biological molecules could be obtained from inorganic precursors. He mimicked the conditions thought to be prevalent in the primitive Earth, and mixed hydrogen (H_2), ammonium (NH_3), and methane (CH_4) in a sealed reactor with an influx of water vapor. Synthesis of a number of organic compounds occurred under the influence of electrical discharges. The de novo synthesized chemicals included amino acids (glycine, alanine, aspartic acid, and glutamic acid), formic, acetic, propionic and fatty acids, cyanide, and formaldehyde (Miller, 1953, 1987). Several researchers followed the Miller's approach using other starting chemical mixes and confirmed that key components of the macromolecules that are associated with living materials (notably purines, pyrimidines, and amino acids) could be made from precursors, which were abundant in the primitive Earth or its atmosphere. Today, variant versions of Miller's protocol (including additional starting chemicals, aerosol spread of chemicals, freeze-thaw cycles, different sources of energy, electron beams, etc.) produce interesting information on the synthesis of organic molecules (Dobson et al., 2000; Miyakawa et al., 2002; Bada and Lazcano, 2003; Ruiz-Mirazo et al., 2014).

Intense ultraviolet (UV) irradiation may have contributed to the synthesis of compounds relevant to life: ammonia, methane, ethane, carbon monoxide, formaldehyde, sugars, nitric acid, and cyanide. Complex organic compounds (notably aromatic hydrocarbons and alcohols) are found in interplanetary dust, comets, asteroids, and meteorites, and they can be generated under the effect of cosmic and stellar radiation. Thus, many organic compounds could have been produced within the Earth atmosphere or away from it, and be transported to the Earth surface by meteorites, comets, or rain, to become the building blocks for additional life-prone organic molecules. Places at which peptide bond formation and prebiotic evolution could have been favored are hydrothermal systems and the interface between the ocean and the atmosphere (Chang, 1994; Horneck and Baumstark-Khan, 2002; Ehrenfreund et al., 2011; Parker et al., 2011; Danger et al., 2012; Griffith and Vaida, 2012; Ritson and Sutherland, 2012; Nakashima et al., 2018).

A key issue from prebiotic syntheses is the degree of oxidation of the primitive Earth atmosphere. Records of an early surface environment dated 3.8 billion years ago were found in metasediments of Isua, Greenland. These materials suggest that the surface temperature of the Earth was below $100°C$, with the presence of liquid and vapor water, and gases supplied by intense volcanism (CO_2, SO_2, and N_2). The composition of primitive rocks, together with theoretical considerations, suggest a neutral redox composition of the Earth atmosphere (with relative gas abundances: N_2, $CO_2 > CO \gg CH_4$, $H_2O \gg H_2$, $SO_2 > H_2S$) around the time of the origin of primeval forms of life. The possible presence of a general or localized reducing atmosphere (N_2, $CO > CH_4 > CO_2$, H_2O, $\sim H_2$, $H_2S > SO_2$) is still debated, but increasingly viewed as unlikely. In an oxidative atmosphere, yields of amino acids, nucleotides, and sugars would be lower. Either these diminished yields were sufficient to attain critical levels of relevant building blocks, or an earlier reducing atmosphere may have accumulated them, among other possibilities (Trail et al., 2011). In spite of the validity of "life comes from life" in the current Earth environment, the experimental facts suggest that there is no barrier for the generation of life from nonlife, provided suitable environmental conditions are met. In this line, A.I. Oparin proposed that a "primitive soup" could well have been the cradle of life on Earth, as described in his famous treatise on the origin of life (Oparin,

1938), a concept that had already been sketched by C. Darwin.

The "protein first" versus "nucleic acid first" for the origin of life is still a contended issue (Falk and Lazcano, 2012). Although during decades there was a preference for nucleic acids due to their superior capacity for self-organization to perpetuate inheritable messages through base pairings, recently both views have accumulated arguments in their favor. An integrative "metabolism first" view based on mutually catalytic networks of small molecules displaying the capacity to replicate and evolve is gaining adepts (Lancet et al., 2018). The building blocks of nucleic acids have been more difficult to obtain from primeval chemicals than the building blocks of proteins. Peptides of about 20 amino acids in length could have been easily formed under prebiotic conditions (Fox and Dose, 1992), and peptides or their derived multimers had a potential to display catalytic activities at a proto-metabolic stage. Peptide amyloids have been proposed as an alternative to nucleic acids for the origin of life in what has been termed *the amyloid-word hypothesis* (Greenwald et al., 2018). Interestingly, amyloid aggregates display features of quasispecies, not in the way of mutant distributions found in viruses (Chapter 3) but as alternative protein conformations and conformation heterogeneity that confers functional diversity (see Chapter 10 on the collective behavior of prions). The protein-versus nucleic acid-first alternatives may not be incompatible. Short nucleotide and amino acid polymers might have cooperated to cross a complexity threshold for self-sustained replication to arise and evolve (Kauffman, 1993). As soon as peptide- or protein-based catalytic activities developed, they had to be coordinated with oligo- or poly-nucleotide replication. This integration of genotypic information with its phenotypic expression may be achieved through hypercyclic couplings, as proposed by M. Eigen and P. Schuster (Eigen and Schuster, 1979; Eigen, 2013), to render incipient life forms sustainable.

1.4.1 Early synthesis of oligonucleotides: a possible ancestral positive selection

Substantiating the abiotic synthesis of nucleotide- or amino acid-based polymers has been arduous relative to the synthesis of monomeric organic molecules. However, early work by L. Orgel and his colleagues documented that polynucleotides could be synthesized from activated nucleotides in the absence of an enzyme (Miller and Orgel, 1974), and recent work has established that there are multiple chemical pathways for abiotic and nonenzymatic nucleotide synthesis (Adamala and Szostak, 2013; Ruiz-Mirazo et al., 2014). Mineral surfaces (clay minerals, zeolites, manganates, hydroxides, etc.) have been proposed as key participants in the origin of life, by providing scaffolds for the synthesis of nucleotide and amino acid polymers (Bernal, 1951; Gedulin and Arrhenius, 1994; Ruiz-Mirazo et al., 2014; Nakashima et al., 2018; Pedreira-Segade et al., 2018)). The relevant features of clays are their adsorption power, ordered structure, capacity to concentrate organic compounds, and ability to serve as polymerization templates. Adsorption onto mineral surfaces may lower the activation energy of intermolecular reactions. Minerals with an excess positive charge on their surfaces might have played a role in the evolution of RNA-like or RNA-precursor molecules. It has been suggested that mineral-organic complexes (chimeras between materials once thought to belong to unmixable categories) could have been the first living organisms endowed with a genetic program (Cairns-Smith, 1965).

Bond formation and chain extension, which are needed to synthesize nucleic acids and proteins prebiotically, are inhibited in liquid water, thus favoring mineral surfaces as potential biogenic sites. Polymer formation could have also been facilitated by heating and drying, with key involvement of the dried phases in catalysis (Towe, 1994; Horneck and Baumstark-Khan, 2002). On different clay types, nucleotides and amino acid polymers of dozens of residues

have been synthesized, providing a realistic scenario for a transition toward primitive self-replicating entities (Ferris et al., 1996). One of the first episodes of Darwinian positive selection could have operated prebiologically through differential surface binding, followed by the selection of autocatalytic heteropolymers (Tkachenko and Maslov, 2018). We arrive at the paradox that while primitive polymerization reactions might have required dry surfaces, later forms of life evolved in a water-rich environment, with water being the major component of cells. "Origin" and "establishment and evolution" of life have been considered either as linked processes that obey similar rules, or distinct events whose investigation requires dissimilar approaches. A distinction between "origin" and "establishment" will become pertinent also when addressing the origin versus the evolution

of viruses later in this chapter, and the mechanisms of viral disease emergence in Chapter 7.

1.4.2 A primitive RNA world

A simplified overview of a course of events that led to the origin of biological systems is depicted in Fig. 1.3. The prebiotic synthesis of potential building blocks—which might have been initiated earlier than 5000 million years ago—renders plausible the existence of a pre-RNA era that was then replaced by an RNA world in the late Hadean early Archean periods on Earth. This stage should have been followed by one in which RNA was complemented by DNA as a repository of genetic information (Bywater, 2012). Polymers other than DNA and RNA are also capable of encoding evolvable inheritable information (Pinheiro et al., 2012;
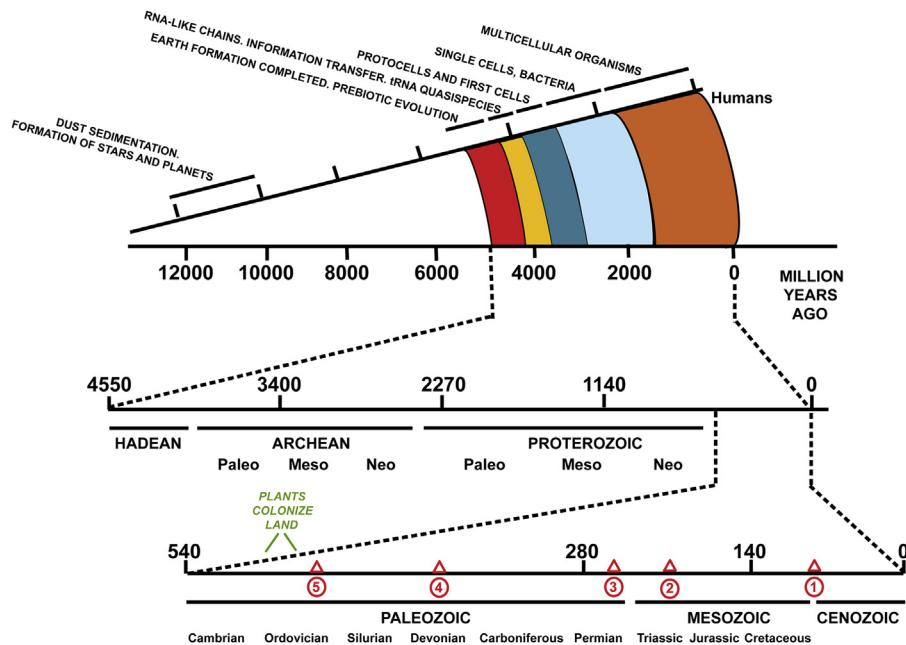

FIGURE 1.3 Schematic representation of geological eons and eras in Earth history, and major prebiological and biological transitions. Time (horizontal lines) is expressed as millions of years before present, counted from the big bang (estimated at ∼ 13,800 million years ago). The time at which plants colonized land (green cone) and major mass extinctions (red triangles) are numbered: 1, end Cretaceous extinction; 2, late Triassic extinction; 3, end Permian extinction; 4, late Devonian extinction; 5, end Ordovician extinction. A proposed sixth extinction presently underway is not indicated. *Illustration by C. Perales and E. Domingo from information retrieved from references given in the text.*

Robertson and Joyce, 2012). Heterogeneous nucleic acid molecules (including mixtures of ribo- and deoxyribo-polymers) can give rise to functional nucleic acids (Gilbert, 1986; Lazcano, 1994a; Gesteland et al., 2006; Derr et al., 2011; Szostak, 2011; Trevino et al., 2011). RNA enzymes (ribozymes), such as RNA ligases can evolve from random-sequence RNAs (Joyce, 2004; Sczepanski and Joyce, 2014). The critical polymerization reaction involves the formation of a phosphodiester bond and release of pyrophosphate—analogous to the reactions catalyzed by the present-day RdRps—and represent an incipient, primitive anabolism (Eigen, 1992; Lazcano, 1994a; Orgel, 2002, 2004; Dworkin et al., 2003; Joshi et al., 2011). In support of a possible link between catalytic RNA activities and solid mineral surfaces in the origin of life (as summarized in Section 1.4.1), the catalytic activity of the hammerhead ribozyme of the Avocado Sun Blotch viroid was maintained when bound to the clay mineral montmorillonite (Biondi et al., 2007).

Minimum requirements for an RNA world would be the presence of ribozymes and mechanisms for the intake of energy-rich molecules (Orgel, 2002, 2004). The inherently low copying fidelity of the putative ribozyme polymerases, estimated in 10^{-2} to 10^{-4} errors per nucleotide copied (Wochner et al., 2011) should have ensured genetic variation for selection to act upon variant RNA molecules. Chirality (the existence of two mirror images or enantiomers of a molecule) poses a challenge for the chemical origin of biological molecules (Caglioti et al., 2011; Ruiz-Mirazo et al., 2014; Sczepanski and Joyce, 2014). Present-day biological systems use only D-ribose (D from "dextro" or rotation of the plane of polarized light to the right) while chemical condensation reactions produce equal amounts of the D- and L- (levo) forms. Nonenzymatic template-dependent reactions can be inhibited by the incorrect enantiomer. This led to the proposal that analogs devoid of enantiomeric forms, such as glycerol derivatives, could

have been the basis of the most primitive genetic systems (Schwartz and Orgel, 1985). Theoretical studies support the notion that initial achiral conditions can evolve toward chirality, in what has been defined as an extension of punctuated equilibrium to prebiological evolution (Gleiser et al., 2008).

Probably, very little, if anything, remains in our present-day biosphere of a primitive RNA world, let alone traces of the prolonged process that went from chemistry to the first replicating organizations, so profound have been the changes experienced by the Earth and its surroundings for over four billion years (Cantine and Fournier, 2018). Contemporary catalytic RNAs (found in the ribosome, as part of some protein complexes, and in some viroids), as well as nucleotide-like coenzymes, have been regarded as possible molecular remnants of a primitive RNA world (Lazcano, 1994a). Some authors consider the possibility that cells whose genetic material is made of RNA (Forterre, 2005) may still hide in some remote sites of our planet (Yarus, 2010). For a transition from a purely RNA (or RNA-like) world to an extended scenario with the participation of proteins, the presence of a transfer RNA (tRNA) quasispecies and the generation of a genetic code endowed with evolutionary potential must have been critical (Eigen, 1992; Koonin and Novozhilov, 2017). Initial theories of how the genetic code might have arisen were put forward by F. Crick, L. Orgel, and C. Woese in the middle of the twentieth century. Main proposals included a stereochemical fit between some amino acids and the corresponding bases (or codons to be), progressive evolution from a one nucleotide to a three nucleotides code, gradual incorporation of amino acids in the coding system, and the frozen accident model of codon universality (for a review of early concepts, see Crick, 1968). New insights on the code origin have come from integrating the knowledge of the mechanisms of protein synthesis with likely events in the RNA world. Primitive tRNA quasispecies (existing around 4000

million years ago, late Hadean, early Archean, Fig. 1.3) and tRNA aminoacylating ribozymes should have evolved to fit the genetic code, which was later expanded in coevolution with the translation machinery (Szathmary, 1999; Rodin et al., 2011; Caetano-Anolles et al., 2013). The age of the genetic code has been estimated in 3.8 ± 0.5 thousand million years (Eigen, 2013), and the present code structure is remarkably redundant. Effects of redundancy are to minimize the deleterious effects of mutations, and to contribute to flexibility of functional involvement of mRNA secondary structure (Shabalina et al., 2006; Koonin and Novozhilov, 2017; see also Chapter 2).

The advent of DNA as an informational macromolecule that was physically more stable than RNA opened the way to an increase of complexity (in the sense of the amount of genetic information) of the genetic material. DNA allowed integration of modules to form the first "chromosomes" and transcriptional regulation prior to protein expression. As in current evolutionary virology, perhaps the most challenging problem to understand early life is to identify the selective constraints that influenced the course of events. In contrast to the present-day environmental changes confronted by viruses (multiple and complex, but amenable to experimentation; Chapters 4 and 6), the conditions that permitted primitive genetic entities to acquire expanded coding and signaling capacities defy our imagination. Some aspects are considered next.

1.4.3 Life from mistakes, information from noninformation: origin of replicons

The most salient attributes of living matter are reproduction, evolvability, energy conversion (metabolism), and compartmentalization. What selective forces might have led to the integration of these features? Concerning reproduction, there must have been a critical transition from the absence of any inheritable instruction

(despite the presence of primitive polymers) to the first molecules endowed with "inheritable" information, for example, a macromolecule capable of self-copying. Such a molecule should have had an immense selective advantage over surrounding macromolecular peers devoid of the capacity for reproduction. The transition from "no inheritable information" to "inheritable information" is essential for the origin of life. Current evidence suggests that the process that allowed such critical transformation was slow and inaccurate. Slow because the catalytic RNAs selected in the laboratory are about 10 million-fold slower than most protein enzymes (Jeffares et al., 1998; Yarus, 2010). Inaccurate because preenzymatic nucleotide polymerization would rarely display error rates below 10^{-1} to 10^{-2} mutations per nucleotide copied (Inoue and Orgel, 1983). No "predators" such as degrading enzymes were present to impede a slow accumulation of variant replicating molecules. The slow sequence exploration might have taken place in microenvironments shielded from damaging radiation until points in sequence space compatible with self-organization, replication, and adaptation were encountered. The adaptive potential of mutant distributions of polynucleotides is a key concept in the quasispecies theory of the origin of life (Eigen and Schuster, 1979), and a signature of present-day viruses (Chapter 3). Replicative inaccuracy and heterogeneity appear as recurrent requirements for the major transitions and adaptability of the forms generated in the course of prebiological and biological evolution. Once a primitive molecular "memory" was implemented, in the words of M. Eigen: " ….information generates itself in feedback loops via replication and selection, the objective being 'to be or not to be'" (Eigen, 1994, 2013). In those times, this was the only simple requirement: to be or not to be.

This singular transition resulted in the first replicating entities that are also termed *replicators or replicons*. They were selected for replicability, stability, and evolvability with trade-offs

(acquisition of benefits for one of the three traits at some cost for another trait) likely play a role at this stage (see Chapter 4 for trade-offs in virus evolution). Optimization of primitive replicons should have been facilitated by dual genotypic and phenotypic features in the same molecule (Eigen, 2013). That RNA itself is both genotype and phenotype is a feature of present-day RNA viruses. That is, the genomic RNA determines phenotypic traits, independently of its protein-coding activity, for example, through its involvement in RNA-RNA and RNA-protein interactions. "Priming" of polynucleotide synthesis, in the sense as we know it today, should not have been a limitation since circular RNA or RNA-like molecules could fold partially to prime their own copying. The term "replicon" currently refers to any genetic element that encodes sufficient information to be copied (i.e., viruses, plasmids, etc.), even if the copying is carried out by (or in conjunction with) elaborate cell-dependent machinery. Virtual replicons are used in computer simulations to learn about the dynamics of natural living systems (Adami, 1998; Eigen, 2013).

The environment in which primitive replicons had to self-organize about 4000 million years ago was very different from the environment we have today on Earth. The sun was about 25% −30% less luminous than the present-day sun, yet it produced more ultraviolet (UV) light. Due to the absence of oxygen and ozone layer, the UV radiation that reached the Earth was 10- to 100-fold more intense than today, with the difference being accentuated for the radiation in the 200−280 nm range. Studies of the conversion of UV radiation into DNA-damage equivalents suggest a two to three logarithm larger biologically relevant UV radiation during the time of the putative RNA world as compared with today's radiation (Canuto et al., 1982; Chang, 1994; Horneck and Baumstark-Khan, 2002). In such an environment, radiation-related mutational input could have had drastic effects on replicating entities in ways that can be only roughly anticipated from the present-day chemistry. Even the simplest present-day RNA genetic systems, despite their small target size, would undergo severe radiation damage. Reconstruction of protein enzyme-free nucleic acid synthesis under the radiation conditions prevalent on Earth during the RNA world development, during late Hadean and early Archean eras, offers a fascinating challenge and opportunity of experimental research for the rising field of Astrobiology.

It has been considered that the time elapsed since the Earth attained a life-friendly environment until protocells arose (from about 4500 million to about 3500 million years ago, Fig. 1.3) was insufficient for life development. This led to the panspermia theory, which proposes that life has an extraterrestrial origin. Panspermia in different forms has been defended by noted scientists, such S. Arrhenius in the early twentieth century and later by F. Crick and L. Orgel (discussed by de Duve, 2002). In addition to the time estimates for life generation being arbitrary, our present understanding of how error-prone replication can facilitate evolvability and exploration of novel biological functions (Chapters 2 and 3) converts a 1 million-year time period in a long time for life to originate and initiate multiple branches for its development.

1.4.4 Uptake of energy and a second primitive positive selection

Energy conversion is an essential feature of life. The primitive precellular organizations might have obtained energy either from organic molecules captured from the external environment (heterotrophy) or from metabolites they synthesized endogenously using external energy (autotrophy). One line of thought considers that it is more likely that the first cells were heterotrophs, and that only later they evolved toward autotrophy, in the form of photosynthesis, which

represented a major transition in the repertoire of biosynthetic pathways (Nakashima et al., 2018). Fermentation reactions were likely the first ones exploited to break energy-rich bonds, as a source of energy for primordial biochemical reactions.

An alternative view is that the first cellular organism was an autotroph, in particular, a chemoautotroph (also termed *lithotroph*) that used inorganic compounds to obtain energy. One of these proposals is that the formation of pyrite from hydrogen sulfide was used by primitive cells as an energy source, resulting from reactions such as $FeS + H_2S \rightarrow FeS_2 + 2H^+ + 2e^-$ (Wächtershäuser, 1994). Positive charges on pyrite crystals could accumulate negatively charged molecules (e.g., the products of CO_2 fixation) and undergo reductive reactions. The system might have selected surface-bound polymers rather than monomers, and given rise to biochirality (selection of one enantiomeric form over another) because of the chiral structure of pyrite (Wächtershäuser, 1988). Then, an evolution toward an Archean carbon-fixation cycle would occur, a precursor of the metabolic cycles found in present-day archaeal and bacterial organisms (Ruiz-Mirazo et al., 2014). The picture may have been more complex, as judged by the success of mixotrophic organisms which are capable of switching from phototrophy (light-mediated break-down of CO_2 for their metabolism) to heterotrophy in some extreme environments of the present-day Earth (Laybourn-Parry and Pearce, 2007).

The integration of early replication systems and metabolism was probably favored by some compartmentalization of replicative-metabolic units through lipid bilayers (Carrara et al., 2012; Stano et al., 2013; Ruiz-Mirazo et al., 2014; Hanczyc and Monnard, 2017). Here a second decisive positive selection event might have entered the scene at the stage of formation of proto-cells and the first individual cells (3600−3200 million years ago, Fig. 1.3; Eigen, 1992). Lipid bilayers endowed with a splitting capacity should have been strongly selected at this stage of life development because of their power to spread. This underlines the concept that Darwinian selection need not be associated exclusively with template-copying processes. Membrane traffic and reorganizations are essential for the life cycle of many present-day viruses (Huotari and Helenius, 2011). Virus particle stability and capacity to spread, associated with membrane capture and interactions, might have derived from the early genomes that exploited membrane-based organelles to achieve functional diversification. Compartmentalization is considered one of the key developments to initiate a cell-based life (Morowitz, 1992; Hanczyc and Monnard, 2017).

1.4.5 Definitions of life

Despite obvious difficulties in reproducing physical and chemical processes that were likely involved in the origin of life, there is sufficient evidence to render probable that the most primitive organizations that we would now consider as "living" resulted from the assembly of simple organic compounds that attained a required level of complexity (Kauffman, 1993). The various facets that distinguish living from inanimate matter are recapitulated in the definitions of life that scientists from different backgrounds in physics, chemistry, or biology have proposed (Box 1.2). Some definitions underline entropy requirements, while others emphasize self-organization, evolvability, metabolism, or regulated complexity. Their disparity reflects the intricacies of a rather enigmatic unfolding of matter.

A. Lazcano summarized the current situation in the arduous search of the origin of life: "There are still unsolved problems but they are not completely shrouded in mystery, and this is no minor scientific achievement. Why should we feel disappointed by our inability to even foresee the possible answers to these luring questions? As the Greek poet Konstantinos Kevafis once

---

BOX 1.2

S o m e   d e f i n i t i o n s   o f   l i f e   a n d   l i v i n g   o r g a n i s m s

- Life is the property of a system that continuously draws negative entropy (maintains orderliness), and delays decay into thermodynamic equilibrium (Schrödinger, 1944).
- Life is an expected, collectively self-organized property of catalytic polymers (Kauffman, 1993).
- Life is a property of any population of entities possessing those properties that are needed if the population is to evolve by natural selection (Maynard Smith and Szathmáry, 1999).
- Life is what is common to all living beings. This answer is not a tautology, as it allows many attributes to be excluded from the definition of life (de Duve, 2002).
- Life is descent with modification. Replication is life's ultimate chemical and physical survival strategy (Yarus, 2010).

- Life is a self-sustained chemical system capable of Darwinian evolution (NASA Astrobiology Institute).
- Living organisms are metabolic-replicating systems composed of molecules and cells which are subject to spontaneous changes in structure and function due to mutations (Demetrius, 2013).
- Life in three statements: (1) Life is not represented by any fundamental physical structure. (2) Life is an overall organization that is governed by functional rather than by structural principles. (3) In order for life to come about, there must exist some physical principle that controls complexity (Eigen, 2013).
- Life, whatever else it may be, is certainly a regularity among material processes (Eigen, 2013).

---

wrote, *Odysseus should be grateful not because he was able to return home, but on what he learned on his way back to Ithaka*. It is the journey that matters" (Lazcano, 1994b; for a general overview of the history of life on Earth, centered in paleontology, see Cowen, 2005).

After this brief survey of the origin of life, we can now examine theories on how, when, and why viruses arose and became active actors in our biosphere.

1.5 Theories of the origins of viruses

Although not in a linear fashion, the number of nucleotides or base pairs in the genetic material—that presumably reflects the amount of genetic information relevant to confer phenotypic traits—increased as evolution led to differentiated organisms. The major theories of the origin of viruses are divided into two opposite categories: those that attribute virus origins to the early development of life, and those that propose that viruses arose when a cellular life was already in place (Fig. 1.4). These two broad views that we may term "viruses *without* cells" and "viruses *from* cells" are not irreconcilable, although reconstruction of ancestral developments is challenging. They can be divided into five main theories—not all independent or mutually exclusive— which are summarized next.
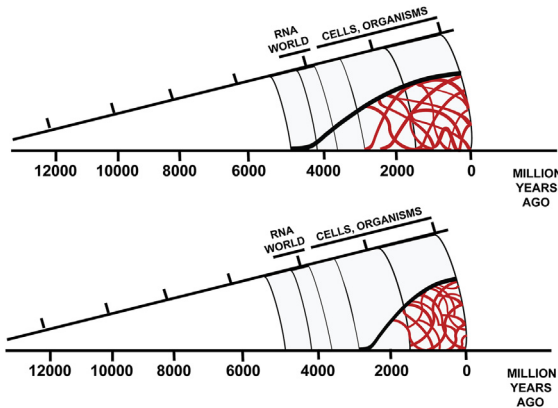
FIGURE 1.4 Two possible courses of events regarding when viruses first appeared and participated in the evolution of the biosphere. The scheme of time frames and major biological events (RNA world, first cells, and organisms) are those displayed in Fig. 1.3. According to the upper diagram, viruses (or previrus-like entities) arose together with the first (precellular) replicating entities. According to the second diagram, viruses (or previrus-like entities) arose when a cellular life had already been established. Presence of virus is generically represented by the external, thick, black curves. The internal red, wavy lines represent generation, dominance, and extinction of multiple viral lineages whose numbers and true dynamics will remain unknown. *Illustration by C. Perales and E. Domingo from information retrieved from the different models of virus origin and references included in the text.*

1.5.1 Viruses are remnants of primeval genetic elements

- Viruses were involved in the origin of life. Some viruses are the descendants of primitive RNA or RNA-like replicons that preceded cellular forms.

Because of their limited genetic complexity, RNA viruses and subviral RNA elements have been considered possible descendants of the primitive replicating entities that predated cell-based life forms (upper diagram in Fig. 1.4). As early as the beginning of the twentieth century, H.J. Muller, L.T. Troland, and J.B.S. Haldane suggested that viruses represented primordial life forms. Influenced by the discovery of bacteriophages by F. d'Herelle, J.B.S. Haldane

proposed viruses as intermediates between the prebiotic soup and primitive cells (reviewed in Lazcano, 2010). In those times, knowledge of viruses was superficial from today's perspective, and lent to daring proposals coherent with viruses being perceived as simple. Despite simplicity and replication being features that could be also be attributed to primitive life forms, we have to distinguish the role that virus-like entities might have played in the establishment of early life from the possibility that present-day viruses reflect how early life might have been.

An origin independent of cells is suggested by the presence of a number of "virus hallmark proteins" in viruses, which do not have counterparts in present-day cells, at least in the sequences represented in data banks. A number of protein folds (such as the double jelly-roll fold) found in the nucleoplasmic large DNA viruses (and related structures) are also present in viruses that infect the archaeon *Sulfolobus*, and in some single-stranded and double-stranded RNA viruses that infect eukaryotic cells. These unique structures and other features of viruses, such as genome size and its composition, constitute a "virus innate self" (Krupovic and Bamford, 2007), that has been taken as an indication that viruses were constructed from modules that were different from those captured to build cells.

The fact that some viral proteins do not have cellular homologs (or are only distantly related to cellular homologs) has favored the view that a virus world preceded the origin of cells (Forterre, 2005; Villarreal, 2005; Koonin et al., 2006; Koonin, 2009). In this line of thought, the present-day RNA viruses are considered remnants of an RNA world; no other biological group that has been identified uses RNA as genetic material. Likewise, transposable elements, repeat sequences of mobile elements (LTR-retrotransposons, LINEs, SINEs, ALUs) and of telomeres and centromeres are considered to be likely of a viral origin [reviewed in Witzany, 2012)].

Other present-day RNA genetic elements that include ribozymes as constituents [i.e., plant viroids, self-replicating RNAs of about 165—430 nucleotides in length (Hadidi et al., 2017) or the defective delta agent, also termed *hepatitis delta virus* (HDV)] might be a vestige of primitive genetic elements (Robertson et al., 1992). HDV is dependent on hepatitis B virus (HBV) for the completion of its infectious cycle (Taylor and Pelchat, 2010). The HDV genome is a mosaic RNA consisting of a viroid-like RNA and an RNA region whose complementary RNA (antigenomic strand) encodes two forms of a structural protein termed *the delta antigen*. Both the genomic and antigenomic RNAs possess a strong secondary structure with about 70% paired nucleotides. The delta antigen is encapsidated by the HBV surface antigen as a component of HDV particles. Thus, HDV appears to be the result of an RNA conjunction between a viroid-like RNA and an mRNA-coding region. Such conjoined RNAs might have been the precursors of the modern eukaryotic genome partition into coding sequences (exons) and intervening sequences (introns) (Sharp, 1985; Robertson, 1992, 1996; Chao, 2007; Taylor and Pelchat, 2010). The structure of the HDV genome seems to echo processes that originated with primitive RNAs selected for their ability to replicate that incorporated a protein-coding moiety through recombination with other RNAs. Replication-competent and protein-coding chimeras might have signaled a relevant intermediate step toward more complex RNA and DNA genomes. The advent of an enzyme that could copy RNA into DNA to carry out reverse transcription should have been instrumental in generating primitive DNA viruses and other DNA-based genomes of increased complexity (Lazcano et al., 1992).

The discovery of the chimeric structure of the HDV genome illustrates how insights into the origin and early evolution of life can be gained from current genomics, despite lacking experimental approaches to recreate episodes that led to virus origins. Even if experiments could be designed, the time frame involved would occupy several generations of scientists, which is not feasible given the current research grant system.

1.5.2 Viruses are the result of regressive microbial evolution

- Viruses originated from regressive evolution of microbes with a cellular organization and became parasites of cells.

This theory is quite the opposite of the previous one because it presupposes that a cellular world was the source of viruses (lower diagram in Fig. 1.4). One of its conceptual pillars is that the intricacies of virus-host relationships unveiled by molecular virology render unlikely that present-day RNA viruses are remnants of an ancestral RNA viral world. Also, the conditions prevailing in the RNA world did not necessitate that a primitive replicon displays rapid replication—a trait of most present-day viruses—for it to become established, because of the scarcity of predators.

The theory of a cellular origin of viruses was already put forward in the twentieth century when it was evidenced that complex DNA viruses encoded enzymes and immunomodulatory proteins that had cellular counterparts. The virus-generating cells could be autonomously functional from the onset (as autonomous as individual cells can be) or belong to a class of simple cells that parasitized functionally more advanced cells.

One of the mechanisms by which viruses may have originated from cells is by regression or reduction. It implies that the number of genes of the virus-precursor cells diminished gradually or in a step-wise fashion to a minimum number compatible with replication at the expense of other cells that supplied basic components for gene expression and energy capture. The giant DNA viruses (Nasir et al., 2012; Colson et al., 2017) have been regarded as a possible missing

link between a primitive DNA cell and the DNA viruses, and some of them include part of the translation apparatus (Abrahao et al., 2018).

The capacity to spread, so inherent to the concept of the virus, might have been first selected as a positive trait for cells, which then regressed toward a subcellular transmissible form. Prokaryotic cells can spread effectively among differentiated eukaryotic hosts, and tumor cells have been regarded as transmissible parasites (Banfield et al., 1965; Murgia et al., 2006; Pearse and Swift, 2006). The observation that some "infectious cells" can be disseminated by insects provides a model for an early origin of arthropod-borne viruses. It has been recognized that transmission of the virus from an infected cell into a recipient cell need not involve a prolonged stay of the virus in the extracellular environment. Infection "synapses" allow intimate cell-to-cell contacts through which virus transmission takes place. It is estimated that synapse-mediated transmission maybe 100-fold more efficient than the transmission of viral particles released into the extracellular environment. This is one of several mechanisms of bloc (multiple particles) transmission, which has several implications for viral quasispecies dynamics (Chapter 3). Acquisition of capacity for long-range transmission in space and time should have provided a sufficient selective advantage for a "cell-to-external environment-to-cell" transmission to coexist with "cell-to-cell" transmission in our biosphere.

The possibility that RNA viruses derive from some "organism" that used RNA as genetic material was suggested initially by D. Baltimore (1980). The evidence that RNA-dependent RNA synthesis is rare in cells suggests that either RNA viruses derived their replicates from a now probably extinct "RNA organism" or that the viral replicates evolved from cellular DNA polymerases (Baltimore, 1980; Forterre, 2005, 2006a,b; Yarus, 2010).

1.5.3 Viruses are liberated autonomous entities

- Viruses originated from cellular DNA or RNA that evolved to embody autonomous replication, and an extracellular step in their replication cycle.

An alternative to the regressive or reduction of virus origin from cells is the escape theory. It proposes that part of the genome complement of cells found a way toward autonomous replication in the sense that they could replicate on their own as parasites of cells. The capacity to replicate could come from the parental cell or be acquired externally. The kind of autonomy attained implied an intracellular phase of their multiplication cycle and an extracellular phase in the form of transmissible particles (virions). This new way of life should have been positively selected if an increased capacity of cell-to-cell transfer conferred an advantage to the cells regarding the acquisition of new traits for functional diversification while maintaining a capacity to sustain virus multiplication.

Decades ago, the view that viruses originated from subcellular organelles was a favored one (see e.g., Joklik, 1974). Despite discernible sequence identity between some viral and mitochondrial DNA sequences, no evidence of viruses having functions encoded in cellular organelles and not in chromosomal DNA has been obtained, perhaps reflecting an earlier relationship between viruses and primitive free-living cells rather than between viruses and modern (eukaryotic) cells (see also Section 1.7)

1.5.4 Viruses are elements for long-term coevolution

- Viruses are as ancient as cells, and coevolved with cells or even with precellular genomic organizations, with which they shared functional modules.

Current genomics of viruses and their host organisms (Bushman, 2002; Mount, 2004; Hacker and Dobrindt, 2006) tends to favor a long history of coevolution between viruses and cells. The structure and functions of many viral proteins do not deviate in any salient way from cellular counterparts. This applies to proteins involved in genome replication, and in proteolytic processing of proteins and polyprotein precursors. In contrast, the "virus innate self" that groups viral proteins and protein folds not shared with cellular counterparts, and that provides a powerful argument for a virus origin independent from cells (Section 1.5.1), does not exclude a coevolutionary mechanism even if all modules were not common to cellular and viral entities over a several million years period (see section 1.7). Comparative genomics suggests that the exchange of functional and structural modules through lateral gene transfers, together with fine adjustments mediated by mutation, have contributed to the coadaptation of cells and autonomous replicons over ancient evolutionary periods (Gorbalenya, 1995; Holland and Domingo, 1998; Jalasvuori and Bamford, 2008; Villarreal, 2008). Even mechanisms that prompt viral variation in cell tropism (Chapter 4) have parallels in differentiated organism As an example, a two amino acids insertion into ectodysplasin—a member of the tumor necrosis-binding family—alters its receptor specificity, and the differential expression of the two protein versions plays a role in epidermal morphogenesis (Yan et al., 2000). Furthermore, as the number of three-dimensional structures for viral cellular enzymes has increased to reach thousands, structural similarities between key cellular and viral enzymes (polymerases, proteases) have become apparent.

Baltimore (1980) proposed that a limited number of "archetypal" proteins could be responsible for RNA virus function. He named as "archetypal" a "positive virus polymerase," a "negative virus polymerase," a set of "surface" proteins, and several "proteases," among other proteins and regulatory elements. The argument that "archetypal" modules could spread among positive- and negative-strand RNA viruses was based on features and mechanisms now recognized as much more profuse than in 1980: the multifunctionality of viral proteins, their capacity to diversify by mutation, and the existence of RNA recombination (Chapter 2). Regulatory strategies were also likely shared by cells and viruses. Small micro-RNAs that now populate the cellular world can act as molecular switches for RNA viral genomes to modulate their replication and gene expression (van Rij and Andino, 2008; Diaz-Toledano et al., 2009). A likely course of events is that the long coevolution of protocells and primitive virus-like elements gradually shaped both cells at the individual and organismal level and the precursors of present-day viruses.

1.5.5 Viruses from vesicles

- "Protoviruses" might have originated in primitive vesicles.

Early protocellular communities probably lacked a cell wall or other compartmentalization barriers, an absence that allowed fluid transfers of metabolites and genetic material (Woese, 2002). The most primitive vesicles based on lipid structures inaugurated the distinction between "internal" and "external" milieu and might have evolved to contain self-replicating macromolecules (Jalasvuori and Bamford, 2008; Adamala and Szostak, 2013; Hanczyc and Monnard, 2017). Vesicles located in favorable microenvironments of a primitive Earth had the potential to exchange small molecules between the "inside" and "outside," and export materials toward other vesicles. Exchanges could assist in the coupling between genome replication and an incipient metabolism.

Depending on their composition, lipid vesicles could form and remain stable at temperatures of about 100°C. Their transfer to lower temperatures might have modulated their permeability prior to

the stage at which peptide or protein transporters were inserted as membrane components. Here again, "heterogeneities" became important: membranes made of mixtures of amphiphiles display increased thermostability, permeability, and tolerance to divalent cations. Budding vesicles endowed with traits such as growth, division, and permeability should have been positively selected for their ability to spread favorable replicating molecules and mediators of protocellular functions. Selected protoviral vesicles became gradually dispensable for the spreading of beneficial genes, and viruses evolved from being solely beneficial entities into also displaying a parasitic behavior that exploited cell resources (compare with Sections 1.7 and 1.8). According to this model, when cells became independent units surrounded by lipid-based membranes, viruses promoted the selection of cells that expressed peptidoglycan molecules on their surface to decrease or prevent virus infection. This transition could mark the onset of an arms race behavior that has been associated with a survival strategy of many present-day viruses. A cell wall allowed vertical transmission of genetic information and rendered the cell metabolically independent. In addition, it provided an osmotic environment suitable for energy production. Genetic information and molecular devices for energy capture, storage, and use were equally important for a sustainable cellular organization. Thus, according to this theory, viruses originated from protoviral elements whose main function was to spread useful genes horizontally. The evolved viruses acted as selective agents to promote microbial evolution and then became established in an increasingly differentiated cellular world (Villarreal, 2005, 2008; Hendrix, 2008; Jalasvuori and Bamford, 2008).

Geological studies indicate that mass extinctions of a brief duration (less than 100,000 years, a short geological time!) occurred at several points when multicellular organisms populated the Earth (some of them indicated in Fig. 1.3).

The end-Permian, end-Triassic, and end-Cretaceous extinctions rank among the most drastic, resulting in profound environmental perturbation (Burgess et al., 2014 and references therein). As depicted in the form of red upward and downward curved lines in Fig. 1.4, perturbations associated with viral extinctions (due to massive host extinctions) and severe bottleneck events have periodically blurred traces of many viruses unknown to us.

Mechanisms implied by each of the five theories summarized here might have participated in the origin of viruses as we know them today, and any model will remain tentative for several obvious reasons. Viruses have not left a fossil record amenable to analysis with current technology. Moreover, viral genomes can evolve at very high rates in response to environmental necessities (Chapter 7), and reconstructions of ancient Earth environments are necessarily imprecise. For these reasons, viruses, independently on when they became active actors in the biosphere (Fig. 1.4), are unlikely to have maintained molecular signatures that could shed light on their remote past.

1.6 Teachings from mycoviruses

The viruses that infect fungi, the mycoviruses, recapitulate key issues on virus origins and diversity (Son et al., 2015; Roossinck, 2019). Two main theories of their origin are that they evolved from plant viruses that invaded fungi or that they have coexisted and coevolved with their host fungi for a long time, in line with the long coevolution model of virus origins (Section 1.5.4). They are diverse regarding the genetic material that can be single-stranded or double-stranded RNA or DNA. They can be either asymptomatic, produce disease, or modify critical host traits, such as virulence, again reflecting a broad range of host interactions observed with

other viral groups. The presence of a dsRNA mycovirus in a fungus is essential for the latter to confer heat tolerance to some plants; these findings have defined a three-way symbiosis required for an important phenotypic trait (Marquez et al., 2007; Roossinck, 2013) (see also Section 1.7.3).

1.7 Being alive versus being part of life

An often asked question is whether viruses are alive or not. The answer is debated, as reflected in the various definitions of virus (Box 1.3). Two opposite views coexist. One of them considers viruses as macromolecular aggregates, that is, viruses are regarded as "chemicals." A second proposal is championed by the "ribovirocell" concept of P. Forterre, which implies that a viable, virus-infected cell contains two different organisms that coexist symbiotically: the cell that produces virus and the replicating virus itself. Another facet of the same duality is manifested when viruses are considered merely as perturbing chemicals

with no participation in the tree of life versus viruses being actually important players in the tree of life. The dual character of viruses as "alive" during intracellular replication and "not alive" outside the cell was stressed in some early literature (e.g., Davis et al., 1968).

As discussed in Section 1.4, if life is best defined as a conglomerate of complementary features, viruses display two of these features: the capacity to replicate and to evolve. Thus, although it is debatable whether viruses qualify as "alive," they are (and most likely have historically been) an integral part of life and the construction of life. Viruses have influenced the tree of life, as we know it, with lateral gene transfers—some mediated by viruses—being a key element in its architecture and differentiation of the different domains of life (Ciccarelli et al., 2006; Weiss et al., 2018). Because of the difficulties of defining "life" unambiguously, the reader might have noticed that in previous sections, the question addressed has been "how is life" rather than "what is life." The question of "why" have viruses persisted in the biosphere is addressed next.

BOX 1.3

Some definitions of virus

- Viruses are strictly intracellular and potentially pathogenic entities with an infectious phase and (1) possessing one type of nucleic acid, (2) multiplying in the form of their genetic material, (3) unable to grow and to undergo binary fission, and (4) devoid of a Lipmann system (Lwoff, 1957).
- Viruses are entities whose genomes are elements of nucleic acid that replicate inside living cells using the cellular synthetic

machinery and causing the synthesis of specialized elements that can transfer the viral genome to other cells (Luria et al., 1978).
- Viruses are replicating microorganisms that are among the smallest of all life forms (first two editions of *Fields Virology*).
- Viruses are transmissible deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) genetic elements that require a cell for multiplication (Domingo and Perales, 2014).

1.7.1 Role of viruses in the evolution of the biosphere

Two general models have been proposed to explain the maintenance of viruses in the biological world:

- Viruses have persisted because they parasitized opportunistically any cellular niche which was compatible with their replication apparatus, and which provided adequate resources.

In this view, viruses are "selfish" replicating elements that became successful when increasingly efficient polymerization activities became part of their life cycles. However, several observations indicate that what was once considered purely "selfish" (also referred to as "junk") cellular DNA may not be that useless after all. Given the intimate connections between cells and viruses, the view that viruses primarily exist because they are mere "selfish" entities seems unlikely. A different issue is that under some particular environments, a virus displays a "selfish" element-like behavior, as one of the outcomes of their having been positively selected. An example, is the blind replication at the expense of elimination of hosts in disease processes, although this is necessarily a transient behavior.

- The presence of viruses was positively selected because they promoted cellular variation and functional diversification.

This proposal relates to some of the theories of virus origins (Section 1.5) and the early evolution of life that implied a role for virus-like entities. According to this view, viruses, together with other subcellular genetic elements (plasmids, retroelements, etc.), penetrated the genetic material of ancestral cell forms, acted as agents of lateral gene transfers, and modified the expression profiles of the recipient cells. Probably, there has been (and there still is) a constant flow of genes between cells and viruses and other mobile genetic elements. The abundance of endogenous retroviruses in the mammalian genomes (García-Montojo et al., 2018) is a clear symptom of such a genetic flow. A nonfunctional viral infectivity factor (Vif) [the HIV-1 protein that can counteract the mutagenic activity of the apolipoprotein B mRNA editing complex (APOBEC) proteins; see Section 2.7 in Chapter 2] was found in the remnant of a rabbit endogenous retrovirus termed *rabbit endogenous lentivirus type K* (Katzourakis et al., 2007). About 8% of the human genome is made of retroviral-like elements. Present-day human endogenous retroviruses probably contribute to pluripotency of human cells (Santoni et al., 2012), and genome regulation (Chuong, 2018). The fusion of cells from the placenta is mediated by syncytin, a protein in HERV-W endogenous retroviruses (Chuong, 2018; Villarreal, 2005).

In addition to the promotion of gene transfers to construct key cellular components, viruses probably acted as selective agents for cells to evolve defense mechanisms against viruses, and this may have originated new cellular functions. Also, viruses could favor the survival of some cell types over others, based on differential cell susceptibility to virus infection, thus contributing to cellular diversification. The need to escape viral infection may have furnished novel cell surface receptor proteins through a selection of cellular escape mutants (Buckling and Rainey, 2002; Saren et al., 2005). Some experimental systems consisting of persistently infected cells in which the cells and the resident virus coevolve (Chapter 6) illustrate how viruses could act as selective agents to promote cellular variation. Such variation would not necessarily involve exchanges of genetic material between the virus and the cells, provided sufficient genetic variation of cells took place.

Multicellular organisms devoid of viral entities should have endured a long-term disadvantage over an alternative scenario with the coexistence of cells and viruses. Selection by viruses need not be restricted to cells, and it can

be extended to entire host organisms and their populations. An abundance of hosts may promote viral epidemiological fitness, which is a factor in viral disease emergence (Chapters 5 and 7). Host subpopulations may be selected by their resistance to epidemic outbreaks by highly pathogenic viruses, such as in the 1918 influenza pandemics or currently with the AIDS or Zika pandemics, and Ebola outbreaks in some parts of Africa. Traditionally, plagues decimated the human population and acted as selective agents for differential survival of individuals.

Selfish-opportunistic and selected-functional are not incompatible models of virus maintenance. Once the instruction to replicate had been positively selected, selfish elements could ensue. As we learn about viral and cellular genomics, the current promiscuity and diversity of viruses (Section 1.3) appear as complementary agencies to promote general biological evolution following Darwinian mechanisms. Viruses might have contributed to the DNA replication machinery of cells, to the formation of the eukaryotic cell nucleus, and to a number of developmental processes (Baranowski et al., 2001; Bushman, 2002; Bacarese-Hamilton et al., 2004; Mallet et al., 2004; Villarreal, 2005, 2008; Forterre, 2006a). Cells are a necessity for viruses and viruses are promoters of cell diversity and as a consequence, of cellular differentiation (compartmentalization and functional specialization).

1.7.2 Current exchanges of genetic material

Present-day viruses reveal several mechanisms of exchange of genetic material that might have roots in early cellular evolution. Temperate bacteriophages (the prototypic example being *E. coli* phage λ) integrate their genomic DNA in the DNA of their host bacteria. The uptake of cellular genes by viruses has been amply documented in transducing bacteriophages (those that can transfer DNA from one bacterium to

another), as well as in RNA and DNA tumor viruses. Even RNA viruses that are not known to include a reverse transcription step in their replication cycle can incorporate host RNA sequences. Replication-competent, cytopathic variants of bovine viral diarrhea virus (a type species of the genus Pestivirus of the important family of pathogens Flaviviridae) can acquire cellular mRNA sequences in their genome, via nonhomologous recombination (Meyers et al., 1989). Insertion of 28S ribosomal RNA sequences into the hemagglutinin gene of influenza virus increased its pathogenicity (Khatchikian et al., 1989). Some defective-interfering particles of Sindbis virus included cellular tRNA sequences at their 5′-ends (Monroe and Schlesinger, 1983). Sequences related to some flaviviruses can persist in an integrated form into the DNA of the insect vectors *Aedes albopictus* and *Aedes aegypti* (Crochu et al., 2004). Endogenous hepatitis B viruses (eHBVs) have been identified in the genomes of birds and land vertebrates (amniotes), crocodilians, snakes, and turtles. The evidence is that eHBVs are more than 207 million years old and that ancient HBV-like viruses infected animals during the Mesozoic Era (Suh et al., 2014, Fig. 1.4). The existence of alternative mechanisms for the integration of viral genetic material into cellular DNA suggests an ancient origin and a selective advantage of exchanges of genetic information in shaping a diverse and adaptable cellular world (Eigen, 1992, 2013; Gibbs et al., 1995; Villarreal, 2005, 2008).

1.7.3 Symbiotic relationships

Symbiosis is an important determinant of coevolutionary interactions between hosts and their parasites (Vorburger and Perlman, 2018). Frequent symbiotic and mutualistic interactions have been established with viruses. Human endogenous retroviruses can protect human tissues and the developing fetus against infection by some exogenous retroviruses (Ryan, 2004).

Some bacteria require bacteriophage to express virulence determinants (Tinsley et al., 2006). Symbiosis can be established between bacteriophages and animals (Barr et al., 2013). Several plant RNA viruses delay the symptoms of abiotic stress, such as those produced by drought and frost (dehydration, osmotic stress, and oxidative stress). Protection is mediated by increased levels of osmoprotectants and antioxidants in the infected plants (Xu et al., 2008).

Symbiotic relationships represent a state of local equilibrium between viruses and hosts, triggered by compatibility and occasionally by mutual benefits. An arms race implied by the virus-host interactions described in previous sections might have been the *modus vivendi* for viruses, only interrupted by occasional armistices. Alternatively, symbiotic and mutualistic interactions might have been the norm, only interrupted by occasional defections by killer personalities that have become the key actors of hospital wards and virology textbooks (Roossinck, 2011; Li and Delwart, 2011). As discussed in connection with natural counterparts of the transition toward error catastrophe in viruses (Chapter 9), cellular editing activities such as those displayed by some of the adenosine deaminase acting on double-stranded RNA and APOBEC proteins are also part of the innate immune response against some viruses. In turn, viruses have evolved multiple functions to counteract the host immune response (Chapter 4). The recruiting of cellular functions to confront viruses attests of transient losses of an equilibrated coexistence. Some middle- and long-term equilibrium between virus and host population numbers must be continuously restored by selective events; otherwise, this book would not have been written.

The evolutionary origin of defense mechanisms against viruses can be regarded as a response to an excessive number of virus-cell interactions. Superinfection exclusion is one of the mechanisms used by present-day cells to limit replication of a virus when another one is actively replicating (or incorporated) into the same cell. Exclusion has a biochemical interpretation in the competition of two viral entities for cellular resources, but it might have been boosted by early cellular adaptation to limit viral invasions (Chapter 4). Likewise, components of the intrinsic and innate immune response that prevent infection and disease might have also been endowed with activities that promoted cell variation for adaptability.

1.8 Virus and disease

The two opposite views of the activity of viruses in the biosphere (i.e., opportunistic occupation of any suitable cellular niche or intimate cooperative coevolution with host cells) would be expected to produce a different proportion of pathogenic viruses. Opportunistic invasions should lead mainly to disease-prone viruses, while long coevolutionary periods should lead to a dominance of nonpathogenic viruses. Not all viruses that have been characterized are pathogenic, and in fact, only a minority of those that exist might be. However, since only a limited number of the viral genomes predicted by metagenomic surveys have been characterized, it is not possible to adventure a proportion of beneficial or neutral versus harmful viruses. In the course of investigations on poliomyelitis, a search for related viruses was undertaken, and a number of new viruses later to be known as *echoviruses* were discovered. They were isolated because they caused cytopathology to cells in culture. The virus-containing samples were from individuals that did not show symptoms of a viral infection. The new isolates were designated as "orphans," meaning viruses without the disease. The term echovirus derives from the enteric cytopathogenic human orphan virus. Some of them, or their close relatives, were later associated with disease syndromes, but others were not. Viruses, as diverse as circoviruses, polyomaviruses, or herpesviruses colonize a

considerable proportion of animals, and only some of the virus types are the direct cause of disease [e.g., postweaning multisystemic wasting syndrome (PMWS) by porcine circovirus type 2, or cancer by some polyomaviruses, among many other examples]. Disease potential is unrelated to viral genome size. PMWS is associated with the smallest mammalian DNA virus genome of only 1.7 Kb, while the almost 100-fold larger herpes virus genomes can coexist with immunocompetent humans without the noticeable disease.

The pathogenic character of a virus depends on the intricacies of virus-host interactions that are poorly understood. The more the knowledge of virus-host interactions, the border between the virus being pathogenic or nonpathogenic becomes fuzzier. Viruses may not damage essential cell functions but may affect dispensable cellular functions. Studies by M.B.A. Oldstone and his colleagues on persistent infections of lymphocytic choriomeningitis virus in neuroblastoma cells demonstrated that the resident virus altered the expression of differentiated cell trait while preserving vital functions (Oldstone et al., 1977). This is one of many examples of virus-induced modifications of the so-called *luxury* functions of cells (Oldstone, 1984). Disease manifestations depend on multiple host and viral factors (allelic forms of several genes, microbiome composition, immune status, coinfections with viruses or other agents, and so on). The dual infections with HIV-1 and hepatitis C virus (HCV) during the end of the 20th century accelerated the evolution of HCV-associated liver disease.

1.9 Viral and cellular dynamics and the tree of life

In the following chapters, molecular mechanisms of virus variation are reviewed in connection with virus survival and host interactions.
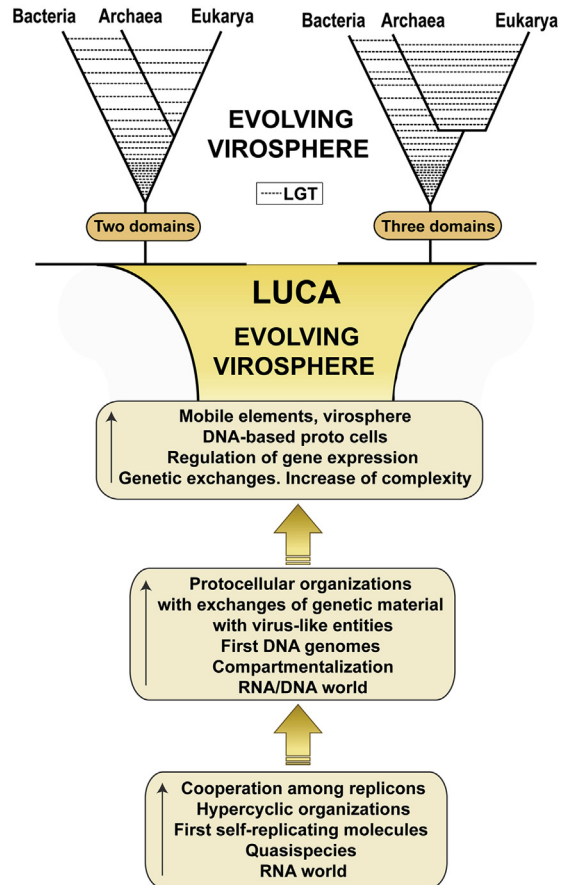

FIGURE 1.5 Scheme of likely occurrences in the origin and early evolution of life and virus involvement, according to current genomics and mechanisms of biological evolution. The time frame involved goes from about 4200 million years ago (bottom) until about 2000 million years ago (top) (compare with geological eras in Fig. 1.3). In each box, the arrow indicates a succession of major developments, probably with overlapping phases. An evolving virosphere is portrayed as having influenced the establishment of the major cellular domains of life. Two of several proposed cell differentiation routes into life domains are drawn. LGT (lateral gene transfer) are depicted as horizontal discontinuous lines, assumed to be more abundant during pre-cellular and early cellular evolution. *Illustration taken from Domingo, E., Perales, C., 2019. Virus evolution. In: Microbial Ecology and Evolution. Encyclopedia of Microbiology. fourth ed. Elsevier, Amsterdam), with permission.*

A reasonable hypothesis is that mechanisms similar to those observed in the present-day biosphere and virosphere were also in operation during the multiple transitions from a primitive replicon-based biosphere to a cellular world differentiated into different domains of life (Fig. 1.5). Although this point will again be examined as part of an overview in the closing chapter of the book (Chapter 10), here, some of the observations on the current biosphere dynamics that bear on the difficulties encountered in reconstructing the sequential transitions toward cellular differentiation are outlined. The following circumstances could have blurred the relationships among evolving previral and precellular entities and among viral genomes under construction, from the perspective of current genomics: (i) episodes of evolution at rates higher than average might historically have accelerated evolution and extinction of viral and cellular genetic elements. (ii) Experimental evolution has documented that viruses often have available multiple mutational pathways to gain fitness in changing environments. This could well apply to other genome assemblies at different stages of complexity. The genome assemblies and linkages we observe are a set among many others possible that could have existed for prolonged time periods but that are now extinct. A related aspect is that the acquisition of genomic modules from other cells or viral elements could render prior acquisitions obsolete or even detrimental. In consequence, an observed absence of shared genomic regions among different viral or cellular lineages from different life domains does not exclude that the shared domains had been transiently (transiently meaning million years!), with no trace in present-day viruses or cells. (iii) Bottleneck events could have erased biological organizations that had a similar degree of adaptation than others that survived. (iv) Fitness variation associated with genomic changes (due to mutation, recombination or lateral gene transfers) is environment-dependent, and the changes in the environment over millions of years might have been dramatic. Such changes might have affected a limited number of local environments or most environments globally. Furthermore, most of these potentially perturbing events might have taken place many times in unpredictable successions, affecting local and few or general and many environments, thus rendering reconstruction of past events from current genomics problematic (Domingo and Perales, 2019). Again, these contingencies will be more apparent when different facets of viral dynamics are dissected in the following chapters.

1.10 Overview and concluding remarks

Contrary to the usual practice in experimental virology, the contents of this chapter have forced a considerable degree of uncertainty, and at times, speculative argumentation. It was, however, a necessary exercise, since at least part of what we see today in viruses must have roots in the origin of life and the role of viruses in life development during epochs that humans have not witnessed. We are left with trying to reconstruct. This chapter is excitingly as close to physics and chemistry as it is to biology. Many of the questions addressed are open to debate, and they will probably remain open for a long time. Interestingly, the terms "heterogeneity" and "complexity" have appeared several times in this chapter, anticipating features of present-day viruses discussed in the upcoming chapters. Complex populations, be they of viruses, cells, protocells, lipid vesicles, primitive replicons, or peptide soups, are the raw materials on which natural selection can act. Primitiveness does not imply simplicity. Complexity was likely a constant trait for early and modern life, both for the extinguished unknown viruses and for those we strive to understand and control (see Summary Box).

Summary box

- Life is characterized by four integrated features: replication, evolvability, metabolism, and compartmentalization. It is debated whether replication or metabolism was the dominant triggering factor.
- At least two ancestral positive selection events might have contributed to the development of life: the selection of replicating over nonreplicating polymers, and selection of splitting-prone versus nonsplitting-prone membrane vesicles.
- Present-day viruses may be descendants from the primitive replicons that participated in early life prior to cellular organizations, or from structured cells by escape or reduction.
- Key features of many present-day viruses may be an inheritance of primitive replicons,

notably error-prone replication, spread through membrane structures, and tendency to engage in genome integration and transfers.
- The geological record indicates multiple, brief, and drastic mass extinction events in Earth's history. Such events anticipate mass extinction of the resident viruses. A dynamics of emergences, reemergences, and extinctions might have operated historically at a grand-scale, with mechanisms similar to those we observe with present-day viruses. Present virosphere dynamics suggests difficulties to reconstruct events that spanned millions of years.

References

Abrahao, J., Silva, L., Silva, L.S., Bou Khalil, J.Y., Rodrigues, R., et al., 2018. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. Nat. Commun. 9, 749.

Abrescia, N.G., Bamford, D.H., Grimes, J.M., Stuart, D.I., 2012. Structure unifies the viral universe. Annu. Rev. Biochem. 81, 795−822.

Adamala, K., Szostak, J.W., 2013. Nonenzymatic template-directed RNA synthesis inside model protocells. Science 342, 1098−1100.

Adami, C., 1998. Introduction to Artificial Life. Springer-Verlag Inc., New York.

Anthony, S.J., Epstein, J.H., Murray, K.A., Navarrete-Macias, I., Zambrana-Torrelio, C.M., et al., 2013. A strategy to estimate unknown viral diversity in mammals. mBio 4, e00598, 13.

Bacarese-Hamilton, T., Mezzasoma, L., Ardizzoni, A., Bistoni, F., Crisanti, A., 2004. Serodiagnosis of infectious diseases with antigen microarrays. J. Appl. Microbiol. 96, 10−17.

Bada, J.L., Lazcano, A., 2003. Perceptions of science. Prebiotic soup—revisiting the Miller experiment. Science 300, 745−746.

Baltimore, D., 1980. Evolution of RNA viruses. Ann. N. Y. Acad. Sci. 354, 492−497.

Banfield, W.G., Woke, P.A., Mackay, C.M., Cooper, H.L., 1965. Mosquito transmission of a reticulum cell sarcoma of hamsters. Science 148, 1239−1240.

Baranowski, E., Ruiz-Jarabo, C.M., Domingo, E., 2001. Evolution of cell recognition by viruses. Science 292, 1102−1105.

Barr, J.J., Auro, R., Furlan, M., Whiteson, K.L., Erb, M.L., et al., 2013. Bacteriophage adhering to mucus provide a non-host-derived immunity. Proc. Natl. Acad. Sci. U.S.A. 110, 10771−10776.

Bengtson, S., 1994. In: Early Life on Earth. Nobel Symposium No. 84. Columbia University Press, New York.

Bernal, J.D., 1951. The Physical Basis of Life. Routledge and Kegan Paul, London.

Bernstein, A.S., 2014. Biological diversity and public health. Annu. Rev. Public Health 35, 153−167.

Berzelius, J.J., 1806. Föreläsningar i Djurkemien, vol. 1−2. Tryckte hos Carl Delén, Stockholm.

Biondi, E., Branciamore, S., Fusi, L., Gago, S., Gallori, E., 2007. Catalytic activity of hammerhead ribozymes in a clay mineral environment: implications for the RNA world. Gene 389, 10−18.

Breitbart, M., Miyake, J.H., Rohwer, F., 2004. Global distribution of nearly identical phage-encoded DNA sequences. FEMS Microbiol. Lett. 236, 249−256.

Buckling, A., Rainey, P.B., 2002. Antagonistic coevolution between a bacterium and a bacteriophage. Proc. Biol. Sci. 269, 931−936.

Burgess, S.D., Bowring, S., Shen, S.Z., 2014. High-precision timeline for Earth's most severe extinction. Proc. Natl. Acad. Sci. U.S.A. 111, 3316–3321.

Bushman, F., 2002. Lateral DNA Transfer. Mechanisms and Consequences. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Bywater, R.P., 2012. On dating stages in prebiotic chemical evolution. Naturwissenschaften 99, 167–176.

Caetano-Anolles, G., Wang, M., Caetano-Anolles, D., 2013. Structural phylogenomics retrodicts the origin of the genetic code and uncovers the evolutionary impact of protein flexibility. PLoS One 8, e72225.

Caglioti, L., Micskei, K., Palyi, G., 2011. First molecules, biological chirality, origin(s) of life. Chirality 23, 65–68.

Cairns-Smith, A.G., 1965. The origin of life and the nature of the primitive gene. J. Theor. Biol. 10, 53–88.

Cantine, M.D., Fournier, G.P., 2018. Environmental adaptation from the origin of life to the last universal common ancestor. Orig. Life Evol. Biosph. 48, 35–54.

Canuto, V.M., Levine, J.S., Augustsson, T.R., Imhoff, C.L., 1982. UV radiation from the young Sun and oxygen and ozone levels in the prebiological palaeoatmosphere. Nature 296, 816–820.

Carrara, P., Stano, P., Luisi, P.L., 2012. Giant vesicles "colonies": a model for primitive cell communities. Chembiochem 13, 1497–1502.

Castón, J.R., Carrascosa, J.L., 2013. The basic architecture of viruses. In: Mateu, M.G. (Ed.), Structure and Physics of Viruses. Springer, Dordrecht, Heidelberg, New York, London, pp. 53–75.

Cello, J., Paul, A.V., Wimmer, E., 2002. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science 297, 1016–1018.

Chang, S., 1994. The planetary setting of prebiotic evolution. In: Bengtson, S. (Ed.), Early Life on Earth. Nobel Symposium No. 84. Columbia University Press, New York, pp. 10–23.

Chao, M., 2007. RNA recombination in hepatitis delta virus: implications regarding the abilities of mammalian RNA polymerases. Virus Res. 127, 208–215.

Chuong, E.B., 2018. The placenta goes viral: retroviruses control gene expression in pregnancy. PLoS Biol. 16, e3000028.

Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., et al., 2006. Toward automatic reconstruction of a highly resolved tree of life. Science 311, 1283–1287.

Colson, P., La Scola, B., Levasseur, A., Caetano-Anollés, G., Raoult, D., 2017. Mimivirus: leading the way in the Discovery of giant viruses of amoebae. Nat. Rev. Microbiol. 15, 243–254.

Cooper, G.M., Temin, R.G., Sugden, B.V., 1995. The DNA Provirus. Howard Temin's Scientific Legacy. ASM Press, Washington, DC.

Cowen, R., 2005. History of Life, fourth ed. Blackwell Publishing, Malden, USA.

Crick, F.H., 1968. The origin of the genetic code. J. Mol. Biol. 38, 367–379.

Crochu, S., Cook, S., Attoui, H., Charrel, R.N., De Chesse, R., et al., 2004. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. J. Gen. Virol. 85, 1971–1980.

Danger, G., Plasson, R., Pascal, R., 2012. Pathways for the formation and evolution of peptides in prebiotic environments. Chem. Soc. Rev. 41, 5416–5429.

Davis, B.D., Dulbecco, R., Eisen, H.N., Ginsberg, H.S., Wood, W.B., 1968. Microbiology. Hoeber Medical Division, Harper and Row, New York.

de Duve, C., 2002. Life Evolving. Molecules, Mind and Meaning. Oxford University Press, Oxford.

Demetrius, L.A., 2013. Boltzmann, Darwin and directionality theory. Phys. Rep. 530, 1–85.

Derr, J., Manapat, M.L., Rajamani, S., Leu, K., Xulvi-Brunet, R., et al., 2011. Prebiotically plausible mechanisms increase compositional diversity of nucleic acid sequences. Nucleic Acids Res. 40, 4711–4722.

Diaz-Toledano, R., Ariza-Mateos, A., Birk, A., Martinez-Garcia, B., Gomez, J., 2009. In vitro characterization of a miR-122-sensitive double-helical switch element in the 5' region of hepatitis C virus RNA. Nucleic Acids Res. 37, 5498–5510.

Dobson, C.M., Ellison, G.B., Tuck, A.F., Vaida, V., 2000. Atmospheric aerosols as prebiotic chemical reactors. Proc. Natl. Acad. Sci. U.S.A. 97, 11864–11868.

Domingo, E., Perales, C., 2014. Virus Evolution. Encyclopedia of Life Sciences. John Wiley & Sons, Ltd, Chichester, UK. https://doi.org/10.1002/9780470015902.a0000436.pub3.

Domingo, E., Perales, C., 2019. Virus evolution. In: Microbial Ecology and Evolution. Encyclopedia of Microbiology, fourth ed. Elsevier, Amsterdam.

Dworkin, J.P., Lazcano, A., Miller, S.L., 2003. The roads to and from the RNA world. J. Theor. Biol. 222, 127–134.

Ehrenfreund, P., Spaans, M., Holm, N.G., 2011. The evolution of organic matter in space. Philos. Transact. A Math. Phys. Eng. Sci. 369, 538–554.

Eigen, M., 1992. Steps Towards Life. Oxford University Press, Oxford.

Eigen, M., 1994. On the origin of biological information. Biophys. Chem. 50, 1.

Eigen, M., 2002. Error catastrophe and antiviral strategy. Proc. Natl. Acad. Sci. U.S.A. 99, 13374–13376.

Eigen, M., 2013. From Strange Simplicity to Complex Familiarity. Oxford University Press, Oxford.

Eigen, M., Schuster, P., 1979. The Hypercycle. A Principle of Natural Self-Organization. Springer, Berlin.

Epstein, J.H., Anthony, S.J., 2017. Viral discovery as a tool for pandemic preparedness. Rev. Sci. Tech. 36, 499–512.

Falk, R., Lazcano, A., 2012. The forgotten dispute: AI Oparin and HJ Muller on the origin of life. Hist. Philos. Life Sci. 34, 373–390.

Ferris, J.P., Hill Jr., A.R., Liu, R., Orgel, L.E., 1996. Synthesis of long prebiotic oligomers on mineral surfaces. Nature 381, 59–61.

Forterre, P., 2005. The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. Biochimie 87, 793–803.

Forterre, P., 2006a. The origin of viruses and their possible roles in major evolutionary transitions. Virus Res. 117, 5–16.

Forterre, P., 2006b. 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.

Fox, S.W., Dose, K., 1992. Molecular Evolution and the Origins of Life. W.H. Freeman, San Francisco.

García-Montojo, M., Doucet-O'Hare, T., Henderson, L., Nath, A., 2018. Human endogenous retroviruses–K (HML-2): a comprehensive review. Crit. Rev. Microbiol. https://doi.org/10.1080/1040841X.2018.1501345.

Gedulin, B., Arrhenius, G., 1994. Sources of geochemical evolution of RNA precursor molecules: the role of phosphate. In: Bengtson, S. (Ed.), Early Life on Earth. Nobel Symposium No. 84. Columbia University Press, New York, pp. 91–106.

Gesteland, R.F., Cech, T.R., Atkins, J.F., 2006. The RNA World. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Gibbs, A., Calisher, C., García-Arenal, F., 1995. Molecular Basis of Virus Evolution. Cambridge University Press, Cambridge.

Gilbert, W., 1986. The RNA world. Nature 319, 618.

Gleiser, M., Thorarinson, J., Walker, S.I., 2008. Punctuated chirality. Orig. Life Evol. Biosph. 38, 499–508.

Gorbalenya, A.E., 1995. Origin of RNA viral genomes; approaching the problem by comparative sequence analysis. In: Gibbs, A., Calisher, C.H., Garcia-Arenal, F. (Eds.), Molecular Basis of Virus Evolution. Cambridge University Press, Cambridge, pp. 49–66.

Greenwald, J., Kwiatkowski, W., Riek, R., 2018. J. Mol. Biol. 430, 3735–3750.

Griffith, E.C., Vaida, V., 2012. In situ observation of peptide bond formation at the water-air interface. Proc. Natl. Acad. Sci. U.S.A. 109, 15697–15701.

Hacker, J., Dobrindt, U., 2006. Pathogenomics: Genome Analysis of Pathogenic Microbes. Wiley-VCH Verlag GmbH&Co KGaA, Weinheim.

Hadidi, A., Flores, R., Randles, J., Palukaitis, P. (Eds.), 2017. Viroids and Satellites. Academic Press, Boston.

Hanczyc, M.M., Monnard, P.-A., 2017. Primordial membranes: more than simple container boundaries. Curr. Opin. Chem. Biol. 40, 78–86.

Hendrix, R.W., 2008. Evolution of dsDNA tailed phages. In: Domingo, E., Parrish, C., Holland, J.J. (Eds.), Origin and Evolution of Viruses, second ed. Elsevier, Oxford, pp. 219–228.

Holland, J.J., 2006. Transitions in understanding of RNA viruses: an historical perspective. Curr. Top. Microbiol. Immunol. 299, 371–401.

Holland, J., Domingo, E., 1998. Origin and evolution of viruses. Virus Genes 16, 13–21.

Horneck, G., Baumstark-Khan, C., 2002. Astrobiology, the Quest for the Conditions of Life. Springer, Berlin.

Huotari, J., Helenius, A., 2011. Endosome maturation. EMBO J. 30, 3481–3500.

Inoue, T., Orgel, L.E., 1983. A nonenzymatic RNA polymerase model. Science 219, 859–862.

Jalasvuori, M., Bamford, J.K., 2008. Structural co-evolution of viruses and cells in the primordial world. Orig. Life Evol. Biosph. 38, 165–181.

Jeffares, D.C., Poole, A.M., Penny, D., 1998. Relics from the RNA world. J. Mol. Evol. 46, 18–36.

Joklik, W.K., 1974. Evolution in viruses. In: Carlile, M.M., Skehel, J.J. (Eds.), Evolution in the Microbial World. Cambridge University Press, Cambridge, pp. 293–320.

Joshi, P.C., Aldersley, M.F., Price, J.D., Zagorevski, D.V., Ferris, J.P., 2011. Progress in studies on the RNA world. Orig. Life Evol. Biosph. 41, 575–579.

Joyce, G.F., 2004. Directed evolution of nucleic acid enzymes. Annu. Rev. Biochem. 73, 791–836.

Katzourakis, A., Tristem, M., Pybus, O.G., Gifford, R.J., 2007. Discovery and analysis of the first endogenous lentivirus. Proc. Natl. Acad. Sci. U.S.A. 104, 6261–6265.

Kauffman, S.A., 1993. The Origins of Order. Self-Organization and Selection in Evolution. Oxford University Press, New York, Oxford.

Khatchikian, D., Orlich, M., Rott, R., 1989. Increased viral pathogenicity after insertion of a 28S ribosomal RNA sequence into the haemagglutinin gene of an influenza virus. Nature 340, 156–157.

Koonin, E.V., Senkevich, T.G., Dolja, V.V., 2006. The ancient Virus World and evolution of cells. Biol. Direct 1, 29.

Koonin, E.V., 2009. On the origin of cells and viruses: primordial virus world scenario. Ann. N. Y. Acad. Sci. 1178, 47–64.

Koonin, E.V., Dolja, V.V., 2013. A virocentric perspective on the evolution of life. Curr. Opin. Virol. 3, 546–557.

Koonin, E.V., Novozhilov, A.S., 2017. Origin and evolution of the universal genetic code. Annu. Rev. Genetics 51, 45–62.

Krupovic, M., Bamford, D.H., 2007. Putative prophages related to lytic tailless marine dsDNA phage PM2 are widespread in the genomes of aquatic bacteria. BMC Genomics 8, 236.

La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., et al., 2008. The virophage as a unique parasite of the giant mimivirus. Nature 455, 100−104.

Lancet, D., Zidovetzki, R., Markovitch, O., 2018. Systems protobiology: origin of life in lipid catalytic networks. J. R. Soc. Interface 15, 20180159.

Laybourn-Parry, J., Pearce, D.A., 2007. The biodiversity and ecology of Antarctic lakes: models for evolution. Philos. Trans. R. Soc. Lond. B Biol. Sci. 362, 2273−2289.

Lazcano, A., 1994a. The RNA world, its predecessors, and its descendants. In: Bengtson, S. (Ed.), Early Life on Earth. Nobel Symposium No. 84. Columbia University Press, New York, pp. 70−80.

Lazcano, A., 1994b. The transition form nonliving to living. In: Bengtson, S. (Ed.), Early Life on Earth. Nobel Symposium No. 84. Columbia University Press, New York, pp. 60−69.

Lazcano, A., 2010. Historical development of origins research. Cold Spring Harb. Perspect. Biol. 2 (11), a002089.

Lazcano, A., Valverde, V., Hernandez, G., Gariglio, P., Fox, G.E., et al., 1992. On the early emergence of reverse transcription: theoretical basis and experimental evidence. J. Mol. Evol. 35, 524−536.

Li, L., Delwart, E., 2011. From orphan virus to pathogen: the path to the clinical lab. Curr. Opin. Virol. 1, 282−288.

Lopez-Bueno, A., Tamames, J., Velazquez, D., Moya, A., Quesada, A., et al., 2009. High diversity of the viral community from an Antarctic lake. Science 326, 858−861.

Luria, S.E., Darnell Jr., J.E., Baltimore, D., Campbell, A., 1978. General Virology, second ed. John Wiley & Sons, New York.

Lwoff, A., 1957. The concept of virus. J. Gen. Microbiol. 17, 239−253.

Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G., et al., 2004. The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. Proc. Natl. Acad. Sci. U.S.A. 101, 1731−1736.

Marquez, L.M., Redman, R.S., Rodriguez, R.J., Roossinck, M.J., 2007. A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. Science 315, 513−515.

Mateu, M.G., 2013. Structure and Physics of Viruses. Springer, Dordrecht, Heidelberg, New York, London.

Maynard Smith, J., Szathmáry, E., 1999. The Origins of Life. From the Birth of Life to the Origins of Language. Oxford University Press, Oxford.

Meyers, G., Rumenapf, T., Thiel, H.J., 1989. Ubiquitin in a togavirus. Nature 341, 491.

Miller, S.L., 1953. A production of amino acids under possible primitive Earth conditions. Science 117, 528−529.

Miller, S.L., 1987. Which organic compounds could have occurred on the prebiotic Earth? Cold Spring Harb. Symp. Quant. Biol. 52, 17−27.

Miller, S.L., Orgel, L.E., 1974. The Origins of Life on Earth. Prentice-Hall, Englewood Cliffs, New Jersey.

Miyakawa, S., Yamanashi, H., Kobayashi, K., Cleaves, H.J., Miller, S.L., 2002. Prebiotic synthesis from CO atmospheres: implications for the origins of life. Proc. Natl. Acad. Sci. U.S.A. 99, 14628−14631.

Monroe, S.S., Schlesinger, S., 1983. RNAs from two independently isolated defective interfering particles of Sindbis virus contain a cellular tRNA sequence at their $5'$ ends. Proc. Natl. Acad. Sci. U.S.A. 80, 3279−3283.

Morowitz, H.J., 1992. Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis. Yale University Press, New Haven and London.

Mount, D.W., 2004. Bioinformatics Sequence and Genome Analysis. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Murgia, C., Pritchard, J.K., Kim, S.Y., Fassati, A., Weiss, R.A., 2006. Clonal origin and evolution of a transmissible cancer. Cell 126, 477−487.

Nakashima, S., Kebukawa, Y., Kitadai, N., Igisu, M., Matsuoka, N., 2018. Geochemistry and the origin of life: from extraterrestrial processes, chemical evolution on Earth, fossilized life's records, to natures of the extant life. Life 8, 39.

Nasir, A., Kim, K.M., Caetano-Anolles, G., 2012. Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea. Bacteria and Eukarya. BMC Evol. Biol. 12, 156.

Oldstone, M.B.A., 1984. Virus can alter cell function without causing cell pathology: disordered function leads to imbalance of homeostasis and disease. In: Notkins, A.L., Oldstone, M.B.A. (Eds.), Concepts in Viral Pathogenesis. Springer-Verlag, New York, pp. 269−276.

Oldstone, M.B., Holmstoen, J., Welsh Jr., R.M., 1977. Alterations of acetylcholine enzymes in neuroblastoma cells persistently infected with lymphocytic choriomeningitis virus. J. Cell. Physiol. 91, 459−472.

Oparin, A.I., 1938. Origin of Life. Dover Publications, New York. English version 1953.

Orgel, L.E., 2002. Is cyanoacetylene prebiotic? Orig. Life Evol. Biosph. 32, 279−281.

Orgel, L.E., 2004. Prebiotic chemistry and the origin of the RNA world. Crit. Rev. Biochem. Mol. Biol. 39, 99−123.

Parker, E.T., Cleaves, H.J., Callahan, M.P., Dworkin, J.P., Glavin, D.P., et al., 2011. Prebiotic synthesis of methionine and other sulfur-containing organic compounds on the primitive Earth: a contemporary reassessment based on an unpublished 1958 Stanley Miller experiment. Orig. Life Evol. Biosph. 41, 201−212.

Pasteur, L., 1861. Memoire sur les corpuscules organizés qui existent dans l'atmosphere. Examen de la doctrine des génerations spontanés. Ann. Sci. Nat. 16, 5−98, 4esue series.

Pearse, A.M., Swift, K., 2006. Allograft theory: transmission of devil facial-tumour disease. Nature 439, 549.

Pedreira-Segade, U., Michot, L.J., Daniel, I., 2018. Effects of salinity on the adsorption of nucleotides onto phyllosilicates. Phys. Chem. Chem. Phys. 20, 1938–1952.

Pinheiro, V.B., Taylor, A.I., Cozens, C., Abramov, M., Renders, M., et al., 2012. Synthetic genetic polymers capable of heredity and evolution. Science 336, 341–344.

Ravantti, J., Bamford, D., Stuart, D.I., 2013. Automatic comparison and classification of protein structures. J. Struct. Biol. 183, 47–56.

Ritson, D., Sutherland, J.D., 2012. Prebiotic synthesis of simple sugars by photoredox systems chemistry. Nat. Chem. 4, 895–899.

Robertson, H.D., 1992. Replication and evolution of viroid-like pathogens. Curr. Top. Microbiol. Immunol. 176, 213–219.

Robertson, H.D., 1996. How did replicating and coding RNAs first get together? Science 274, 66–67.

Robertson, M.P., Joyce, G.F., 2012. The origins of the RNA world. Cold Spring Harb. Perspect. Biol. 4 (5), a003608.

Robertson, B.H., Jansen, R.W., Khanna, B., Totsuka, A., Nainan, O.V., et al., 1992. Genetic relatedness of hepatitis A virus strains recovered from different geographical regions. J. Gen. Virol. 73, 1365–1377.

Rodin, A.S., Szathmary, E., Rodin, S.N., 2011. On origin of genetic code and tRNA before translation. Biol. Direct 6, 14.

Rohlfing, D.L., Oparin, A.I., 1972. Molecular Evolution: Prebiological and Biological. Plenum Press, New York and London.

Roossinck, M.J., 2011. The good viruses: viral mutualistic symbioses. Nat. Rev. Microbiol. 9, 99–108.

Roossinck, M.J., 2013. Plant virus ecology. PLoS Pathog. 9, e1003304.

Roossinck, M.J., 2019. Evolutionary and ecological links between plant and fungal viruses. New Phytol. 221, 86–92.

Ruiz-Mirazo, K., Briones, C., de la Escosura, A., 2014. Prebiotic systems chemistry: new perspectives for the origins of life. Chem. Rev. 114, 285–366.

Ryan, F.P., 2004. Human endogenous retroviruses in health and disease: a symbiotic perspective. J. R. Soc. Med. 97, 560–565.

Santoni, F.A., Guerra, J., Luban, J., 2012. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. Retrovirology 9, 111.

Saren, A.M., Ravantti, J.J., Benson, S.D., Burnett, R.M., Paulin, L., et al., 2005. A snapshot of viral evolution from genome analysis of the tectiviridae family. J. Mol. Biol. 350, 427–440.

Schrödinger, E., 1944. What is Life? Cambridge University Press, Cambridge.

Schwartz, A.W., Orgel, L.E., 1985. Template-directed synthesis of novel, nucleic acid-like structures. Science 228, 585–587.

Sczepanski, J.T., Joyce, G.F., 2014. A cross-chiral RNA polymerase ribozyme. Nature 515, 440–442.

Selander, R.K., 1976. Genetic variation in natural populations. In: Ayala, F.J. (Ed.), Molecular Evolution. Sinauer Associates Inc., Sunderland, Massachusetts, pp. 21–45.

Shabalina, S.A., Ogurtsov, A.Y., Spiridonov, N.A., 2006. A periodic pattern of mRNA secondary structure created by the genetic code. Nucleic Acids Res. 34, 2428–2437.

Sharp, P.A., 1985. On the origin of RNA splicing and introns. Cell 42, 397–400.

Son, M., Yu, J., Kim, K.H., 2015. Five questions about mycoviruses. PLoS Pathog. 11, e1005172.

Stano, P., D'Aguanno, E., Bolz, J., Fahr, A., Luisi, P.L., 2013. A remarkable self-organization process as the origin of primitive functional cells. Angew. Chem. Int. Ed. Engl. 52, 13397–13400.

Suh, A., Weber, C.C., Kehlmaier, C., Braun, E.L., Green, R.E., et al., 2014. Early mesozoic coexistence of amniotes and hepadnaviridae. PLoS Genet. 10, e1004559.

Suttle, C.A., 2007. Marine viruses—major players in the global ecosystem. Nat. Rev. Microbiol. 5, 801–812.

Szathmary, E., 1999. The origin of the genetic code: amino acids as cofactors in an RNA world. Trends Genet. 15, 223–229.

Szostak, J.W., 2011. An optimal degree of physical and chemical heterogeneity for the origin of life? Philos. Trans. R. Soc. Lond. B Biol. Sci. 366, 2894–2901.

Taylor, J., Pelchat, M., 2010. Origin of hepatitis delta virus. Future Microbiol. 5, 393–402.

Tinsley, C.R., Bille, E., Nassif, X., 2006. Bacteriophages and pathogenicity: more than just providing a toxin? Microb. Infect. 8, 1365–1371.

Tkachenko, A.V., Maslov, S., 2018. Onset of natural selection in populations of autocatalytic heteropolymers. J. Chem. Phys. 149, 134901.

Towe, K.M., 1994. Earth's early atmosphere: constraints and opportunities for early evolution. In: Bengton, S. (Ed.), Early Life on Earth. Nobel Symposium No. 84. Columbia University press, New York, pp. 36–47.

Trail, D., Watson, E.B., Tailby, N.D., 2011. The oxidation state of Hadean magmas and implications for early Earth's atmosphere. Nature 480, 79–82.

Trevino, S.G., Zhang, N., Elenko, M.P., Luptak, A., Szostak, J.W., 2011. Evolution of functional nucleic acids in the presence of nonheritable backbone heterogeneity. Proc. Natl. Acad. Sci. U.S.A. 108, 13492–13497.

van Rij, R.P., Andino, R., 2008. The complex interactions of viruses and the RNAi machinery: a driving force in viral evolution. In: Domingo, E., Parrish, C.R., Holland, J.J. (Eds.), Origin and Evolution of Viruses, second ed. Elsevier, Oxford, pp. 161–182.

Villarreal, L.P., 2005. Viruses and the Evolution of Life. ASM Press, Washington, DC.

Villarreal, L.P., 2008. The widespread evolutionary significance of viruses. In: Domingo, E., Parrish, C.R., Holland, J.J. (Eds.), Origin and Evolution of Viruses, second ed. Elsevier, Oxford, pp. 477−516.

Vorburger, C., Perlman, S.J., 2018. The role of defensive symbionts in host-parasite coevolution. Biol. Rev. Camb. Philos. Soc. 93, 1747−1764.

Wächtershäuser, G., 1988. Before enzymes and templates: theory of surface metabolism. Microbiol. Rev. 52, 452−484.

Wächtershäuser, G., 1994. Vitalists and virulists: a theory of self-expanding reproduction. In: Bengtson, S. (Ed.), Early Life on Earth. Columbia University Press, New York, pp. 124−132.

Weiss, M.C., Preiner, M., Xavier, J.C., Zimorski, V., Martin, W.F., 2018. The last universal common ancestor between ancient Earth chemistry and the onset of genetics. PLoS Genet. 14, e1007518.

Whon, T.W., Kim, M.S., Roh, S.W., Shin, N.R., Lee, H.W., et al., 2012. Metagenomic characterization of airborne viral DNA diversity in the near-surface atmosphere. J. Virol. 86, 8221−8231.

Witzany, G., 2012. From molecular entities to competent agents: viral-infection-derived consortia act as natural genetic engineers. In: Witzany, G. (Ed.), Viruses: Essential Agents of Life. Springer + Science Business Media, Dordrecht, pp. 407−419.

Wochner, A., Attwater, J., Coulson, A., Holliger, P., 2011. Ribozyme-catalyzed transcription of an active ribozyme. Science 332, 209−212.

Woese, C.R., 2002. On the evolution of cells. Proc. Natl. Acad. Sci. U.S.A. 99, 8742−8747.

Xu, P., Chen, F., Mannas, J.P., Feldman, T., Sumner, L.W., et al., 2008. Virus infection improves drought tolerance. New Phytol. 180, 911−921.

Yan, M., Wang, L.C., Hymowitz, S.G., Schilbach, S., Lee, J., et al., 2000. Two-amino acid molecular switch in an epithelial morphogen that regulates binding to two distinct receptors. Science 290, 523−527.

Yarus, M., 2010. Life from an RNA world. The Ancestor Within. Harvard University Press, Cambridge, Massachusetts and London, England.