Sankar Chatterjee

From Stardust to First Cells

The Origin and Evolution of Early Life

Springer

Sankar Chatterjee

From Stardust to First Cells

The Origin and Evolution of Early Life

Springer

Sankar Chatterjee
Department of Geosciences
Museum of Texas Tech University
Lubbock, TX, USA

ISBN 978-3-031-23396-8 ISBN 978-3-031-23397-5 (eBook)
https://doi.org/10.1007/978-3-031-23397-5

Cover Illustration: An artistic concept of the early Earth-Moon system about four billion years ago when first life emerged. The Moon would have been close enough to have filled much of the night sky. Meteorite impacts created innumerable craters on the ancient supercontinent Vaalbara like the surface of the Moon. These terrestrial craters filled with water formed hydrothermal crater lakes. Meteorite bombardments delivered life's raw material to young Earth, accumulating in these terrestrial hydrothermal crater vents and paving the way for life.

Artwork: La Gina Fairbetter. Jacket Design: Kendra Dean Wallace

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To my younger son Shuvo—a brilliant engineer—for upgrading my computer skills

Preface

The origin of life from non-life on Earth remains an enduring mystery. Its resolution would not only slake our human need to understand this most central of existential concerns but would guide us in our long and expanding search for life beyond Earth. It is challenging to study life's origin on Earth because it began at least four billion years ago and only left tantalizing geological traces in the form of microfossils and stromatolites in the oldest rocks. A record of the prebiotic processes is not preserved. However, it seems likely that life on Earth was not the outcome of an isolated event but a cosmic imperative and planetary property that appeared on young Earth. Speculation regarding the precise historical path from inanimate to animate—the identity of the building blocks of life available at particular physical locations on prebiotic Earth together with the chemical structures of possible intermediate stages on the road to life—is being rapidly refined using techniques and data arising from many different scientific disciples. This book chronicles the recent breakthroughs in astrobiology that have contributed to understanding the origin of life from a cosmic perspective.

How life created so much diversity from so little physical matter is remarkable. Darwin immortalized this view of life in the last sentence of *The Origin of Species*: 'from so simple a beginning, endless forms most beautiful and most wonderful have been and are being evolved.' The variety of life on Earth is immense and wondrous. Yet, despite their endless variations, the cells that make up all living things contain just three fundamental components in a symbiotic relationship. Nucleic acids such as RNA and DNA are molecules that encode digital information and can be copied. Proteins are the catalysts and nanobots that perform crucial tasks for cells. Encapsulating all of this is a plasma membrane made from a semipermeable lipid bilayer that protects the cell's interior from the outside environment. The intrinsic properties of cells do not come from individual cell components but their collaborative dynamics mediated by sophisticated information systems. In protocells—the precursors of all life—the beginnings of this trinity of molecules appeared stepwise, a collection of self-organized RNA and proteins enclosed by a simple membrane. Side by side, the information system coevolved with the molecules of life. The end product of a long biochemical voyage was the emergence of the first cells capable of identical division. The emergence of these first cells was, quite possibly, the most momentous event in the history of our planet, transforming a sterile world into a living *Gaia*. This book attempts to reconstruct this long molecular voyage by bringing together new insights from astrobiology regarding how heat-loving microbial life could have started on young Earth.

Early in Earth's history, both volcanism and impact cratering were ubiquitous processes. Both may have created an environments to transform disparate chemical compounds into living organisms, providing a suitable habitat for life to evolve. I will begin my narrative with this violent Hadean world, capable of taking the building blocks of life with embedded analog information from meteorite impacts and sparking life. I will discuss the chemical evolution of biomolecules in hydrothermal impact crater basins, the most likely birthplace of life. I will discuss the algorithmic origin of life and how viruses might have played a critical role in prebiotic synthesis and the emergence of the first cells. I will show why there has been renewed optimism in recent years about the existence of life beyond Earth.

This book seeks to explain my view of the present state of the origin-of-life problem. It will be a compendium for curious minds. Its aim is a much wider audience than the science community. This book is very much the product of personal research and reflection on the origin of life in my later years, which served to rekindle my childhood fascination with life, planets, and stars. I spent a lifetime thinking about these questions as an evolutionary biologist and geologist, and I now find myself increasingly drawn from the younger field of astrobiology that is amid rapid advances.

Objectives

Traditionally, there are three groups of scientists dominating the search for the Origin of Life: the organic chemists (chemical evolution), the inorganic chemists (metabolism, energy), and the molecular biologists (RNA world), all of which have based heavily on laboratory experiments. However, life is a planetary phenomenon. Without the knowledge of early Earth and its birth from supernova explosion, our search for the origin of life remains incomplete. To remedy this deficiency, a new interdisciplinary field astrobiology has emerged recently that asks profound scientific questions in the cosmic context. What is the definition of life? What are the sources of the building blocks of life? Where did life begin? When did life begin? What was the energy for prebiotic synthesis? How did life originate and evolve? What is the affinity of early life on Earth? Is there any life elsewhere in the Universe?

This book is the first comprehensive account of the origin of life based on astrobiology. Where does the quest for the origin of life stand today? As this book shows, quite a lot is known, at least in broad outline. So even though detailed understanding has yet to be achieved, the main story of life's beginning is emerging from a recent stride in astrobiology. Thus, the narrative in this book begins not on Earth but in the interstellar medium during a supernova explosion. Although the seeds of life are present everywhere in the interplanetary dust and meteorites, life is exclusively a planetary phenomenon, so far only known in our Goldilocks planet. What was there on young Earth that sparked life? I want to answer these questions in my book based on the recent achievements in astrobiology research and thinking.

The unique feature of this book is the emphasis on the fundamental components of life—biological information systems—that are totally absent in the inorganic world. The book is based on the *information paradigm*, the idea that 'life is chemistry plus information.' Biological information systems exist in three forms: analog, hybrid, and digital. It is clearly incorrect to suppose that biological systems are encoded digitally in DNA alone, because a daughter cell does not only inherit its DNA from its parent cell. It also inherits its translation machine (hybrid information), and cytoplasm and cell membrane (analog information). Throughout the book, I have documented how biomolecules and three classes of information systems coevolved during the origin of life. I have discussed extensively on the role of virus in the origin of DNA and critical enzymes for its transcription and replication, another neglected topic in the origin of life book.

This book has 20 chapters ordered logically and chronologically from simple to complex biomolecules in hierarchical fashion; each chapter finishes with a conclusion and references. The chronological story of the origin of life has a certain rhythm and emerging complexity from a simple beginning. I start with a historical understanding of life's beginnings and present two of the major theories for the origin of life, like panspermia and the origin of life. Next, I describe the early attempts for simulation of prebiotic Earth conditions like the famous Miller-Urey experiment and the first asteroid analysis. The book logically continues with the definition of life and the complexity of the main features of life. Chapter 3 describes biological information, which is highly neglected in the origin of life research. Information is one of the key attributes of life, but the origin of prebiotic information remains a mystery. I show biological information systems exist in three forms: analog, hybrid, and digital. Life's biological and informational evolution are intertwined and inseparable. Life and its information systems form

a closely coupled entity, influencing each other in a complex feedback loop. Information streamlined the prebiotic synthesis from chaotic molecular assemblages and provided directionality.

In Chap. 4, I describe the main chronological events in five hierarchical stages of increasing molecular complexity: the cosmic stage, the geological stage, the chemical stage, the digital information stage, and the biological stage. In the cosmic stage, I describe our cosmic origin, when a supernova explosion nearby the solar nebula cast the building blocks of life into interstellar space. During the Late Heavy Bombardment period, the comets and carbonaceous asteroids transported water and organic molecule to young Earth. Asteroid collisions created numerous hydrothermal crater lakes on the Eoarchean crust, crafting cradles for prebiotic chemistry.

In the geologic stage, crater basins containing an assortment of cosmic and terrestrial organic compounds powered by hydrothermal and chemical energies drove the early processes of prebiotic synthesis. An analog information system (AIS) emerged to move and concentrate building blocks in the vent environment. Here I portray the image of young Earth where genesis began from cosmic building blocks of life. I reconstruct the Eoarchean continental crust with enclaves of sediments in early Archean greenstone basins before the onset of plate tectonics. These hydrothermal sediments in the greenstone facies of the oldest cratons are probably the most likely prebiotic environment where life began on bleak, lifeless rocky protocontinents, surrounded by global ocean water. The oldest record of life, nearly four-billion-year-old-sediments, supports this scenario.

The next chapter describes hyperthermophiles with hydrothermal systems (submarine black and white smokers), which give the reader a good background of the likely cradles of life. Here I describe the possible locations of origin of life on Earth, such as hydrothermal vents or surfaces on different craters with variable wet and dry cycles conducive to condensation reactions in prebiotic synthesis.

With the environmental stage set for the origin of life, Chap. 7 is focused on the primitive metabolism, thermodynamics, and energy sources required for abiogenesis. As a logical progression, the book continues with the prebiotic chemistry and formation of the amino acids, nucleotides, and lipids through a polymerization process at the pores of mineral surfaces on the floor of the hydrothermal crater basins, giving rise to peptide and RNA polymers. At that moment, I begin my presentation by introducing the RNA world and the peptide/RNA world as two major competing theories for the origin of life. I describe in detail the specificity of RNA as polymers and the ability to perform catalytic activity as a ribozyme. Also, I present different variants for RNA self-replication. Here I criticize and show the disadvantages of the lonely RNA world as a probable solution to the origin of life. On the other hand, the most common reaction products from many prebiotic syntheses of small molecules are amino acids. As a likely alternative, I recognize the peptide/RNA world as a better solution. We attempt to understand the potential biochemical pathway for abiogenesis.

In the peptide/RNA world, several noncoding RNA molecules such as pre-tRNA, activating enzymes such as bridge peptide and pre-aaRS, and ribosomes became the ancestral components of the translation machine and ushered the hybrid information system (HIS). These hybrid components build the translation machinery step by step. Here I describe the steps for the development of the components of the translation machine.

HIS gave rise to the formation of a digital information system, which is central to the origin of life. In the first step, amino acids were selected that would play a pivotal role in translation and genetic code. In the next step, tRNA molecules become molecular architects for designing and assembling mRNA step by step, employing their two distinct genetic codes. First, they created codons of mRNA by the base pair interaction (anticodon-codon mapping). Secondly, each charged tRNA transferred its amino acid information to the corresponding codon (codon-amino acid mapping), facilitated by an aaRS enzyme. This is the moment of the emergence of the first gene before DNA.

In the next chapter, there is a long discussion about the discovery and specificity of the genetic code, the following with the hypothesis of the genetic code. Here I offer a new view of the coevolution of translation machines and the genetic code. With the advent of encoded mRNA, a mapping mechanism was developed between each codon and its cognate amino acid. As more and more codons 'remembered' their respective amino acids, this mapping system developed the genetic code in their 'memory bank.' With the genetic code in place, a unidirectional flow of information originated from mRNA to protein.

Chapter 14 describes the evolution of proteins by translating encoded mRNA associated with the return of an analog information system. The specificity of the protein enzymes, the central enzyme today, is discussed in detail. The formation of and the origin of phospholipid membranes are the following paragraphs. The chapter also suggests how the protein/RNA world might have given rise to the first virus.

The newly synthesized proteins from the nucleotide sequences of mRNA led to the emergence of RNA viruses. Although RNA virus was a hybrid molecule of RNA and protein, it ushered in a new digital information age by horizontal gene transfer. As RNA viruses infected RNA protocells, retrovirus evolved. Retrovirus, in turn, created DNA and donated DNA and important enzymes for transcription and translation to protocells during recurrent infection. DNA in protocells became the primary genetic storage system. Information began to flow from DNA to RNA to proteins when the central dogma of molecular biology was born.

The critical breakthrough to life arrived in the biological stage with the emergence of the first cells capable of self-production, reproduction, heredity, variation, and Darwinian evolution. The origin of the first cells was an event horizon that transformed the rocky, sterile world into a living, evolving world. Here I discuss LUCA as the hypothetical hyperthermophilic ancestor of all life that created two domains of life, Bacteria and Archaea, in the Archean. The next chapter discusses the habitat and nature of Archean life from extensive fossil evidence.

In Chap. 19, I describe the possibilities of life in exoplanets in the solar system and elsewhere in the Universe. I consider Extraterrestrial Intelligence (SETI) and whether there is any other intelligence in the Universe with which we can communicate. We contemplate the future and fate of our own civilization. In the final epilogue of the book, I summarize the salient points in the origin of life: how chaotic assemblages of biomolecules led to the highly ordered life systems, the emergence of hierarchy, informations systems, complexity, and directionality in the prebiotic synthesis, entropy and life, role of viruses in the origin of life, and life in extreme environments.

Lubbock, TX, USA Sankar Chatterjee

Acknowledgments

I have significantly benefited from conversations and correspondences with several colleagues and friends. I owe an immense debt of gratitude to the many authors whose studies have helped me with their thoughtful, well-documented, and enlightening expositions. I especially thank David Deamer, who inspired me in my journey into astrobiology. The stimulus for writing this book is that scientists will someday discover how life can emerge on habitable planets like Earth and Mars.

I have had the good fortune to receive help from many colleagues and friends in writing this book. I am grateful to Francis Crick, Freeman Dyson, Christian de Duve, David Deamer, Eugene Koonin, Paul Davies, Ray Kurzweil, Kenzi Ikehara, Robert Hazen, Steven Benner, and Lynn Margulis for their insights, which have been incorporated and expanded upon in this book. I thank Lynn Margulis, David Deamer, Norman Sleep, Bill Glen, Bruce Clark, Michael San Francisco, and Surya Yadav for helpful discussions and constructive suggestions at different stages. I thank Surya Yadav for collaboration on biological information systems. I owe an immense debt of gratitude to the many authors whose basic research has helped shape our understanding of life's origin. I thank David Deamer, Sam Braudt, Jacob Bowe, and Bruce Clark for their careful edits of the manuscript and helpful suggestions. I thank Volkan Sarigul, Oliver McRae, and Jeff Martz for the beautiful artwork. La Gina Fairbetter created an attractive book cover. The world of molecules is completely invisible. Volkan created the illustrations in this book to help bridge the gulf and allow us to see the molecular structure of protocells and the birth of the first cells and gave a lovely pictorial rendition to my text. I thank Simone Marchi for permission to use Figs. 4.1 and 5.1. A P. W. Horn Distinguished Professor grant from Texas Tech University and a Fulbright-Nehru Excellence Award supported this work. I must thank NASA and the NASA Astrobiology Institute for the valuable information and several images I have used in this book.

Working with the professionals of Springer Nature has been a thoroughly enjoyable and rewarding experience. I thank Ron Doering, the publishing editor, for inviting me to write this book, Murugesan Tamilsevan, the Project Coordinator, for waiting patiently for several years to get the manuscript and aided me throughout the preparation of the manuscript with endurance and encouragement, and Sanjana Sundaram for leading this manuscript to completion. Finally, Zachary Romano, the Senior Editor of Springer, has contributed to every stage of its development, assisted by Zoe Kennedy. My sincere thanks go to all the Press's editorial staff for improving the style and clarity of my writing. Texas Tech University has provided an excellent academic environment and continued funding my research for four decades. I want to thank Jamie Looney and Kendra Dean Wallace of the TTU museum for their continued technical help. This book recognizes Texas Tech University's centennial in 2023.

Finally, I wish to thank my two sons, Soumya and Shuvo, for stimulating discussions on science, religion, music, and politics during their visits that have kept my mind sharp and active in my old age. Shuvo provided critical information on machine learning. My wife, Sibani, can speak to how much she has endured while I was struggling with this project in my study alone. I thank her for her love and support.

Lubbock, TX, USA Sankar Chatterjee

Contents

Prologue

1

In the beginning, there was only darkness, veiled in darkness.
In profound darkness, water without light.
All that existed then was void and formless.
That one arose, at last, born of the power of heat.

—Creation Hymn, The Rig Veda, 1500 BC, India

1.1 Genesis

From space, Earth is a dazzling blue jewel delicately wrapped in a mist of the biosphere. On December 7, 1972, when the astronauts of the 'Apollo 17' spacecraft circled Earth in their tiny capsule, millions of listeners heard them describe the beauty of this planet, likening it to a 'Blue Marble,' in a moment of extraordinary human revelation. The National Aeronautics and Space Administration (NASA) subsequently captured a 'Blue Marble' image of Earthrise as viewed from the Moon at a much higher resolution (Fig. 1.1). It is one of the most epochal photographs of not just ours but of all time. It depicts the human condition living alone on a stunning island ecosystem amidst a vast and unknown universe.

As far as we know, the presence of life makes our planet unique. Life on this planet provokes perpetual awe and wonder over this cosmic drama. However, some four and a half billion years ago, when Earth formed from a spinning disk of cold, diffuse cloud and dust swirling around the Sun, it was a lifeless, waterless, airless, barren planet. Half a billion years later, Earth evolved from a molten rocky mass to a living world with an atmosphere, protocontinents, and vast oceans teeming with tiny microbes. What was it in that violent sky and primordial waters that turned organic molecules into organisms? What triggered the origin of life on virgin Earth? How exactly did a lifeless planet—prebiotic Earth—become a wonderland of diverse life forms? The eternal question of how life on Earth began and evolved has intrigued all civilizations. What essential differences allow a living thing to grow and reproduce but not a stone? Strictly speaking, the question of the emergence of life is not only a scientific inquiry but also an existential one—a historical and philosophical challenge to human comprehension of the world. The search for this answer is the story of our quest to discover our ultimate origin, the riddle of the birth of life on Earth. Every culture has a creation myth; every generation must rewrite the *Book of Genesis*.

The origin of life is one of the great mysteries of science. We now know, in broad strokes, how life, once it began, proceeded to multiply and diversify by Darwinian evolution until it filled every niche on the planet. However, what is the origin of life itself? How did life begin on planet Earth? How did we get from a sterile planet to one with a rich abundance of life that we see today? Darwin pondered this question but had no answer. The origin and diversity of life have baffled humans for millennia. In this perennial quest, many different processes have been proposed, all of them controversial. On the one hand, Leslie Orgel, a highly regarded origin-of-life chemist, nicely summarized this difficulty in 1998:

> The problem of the origin of life on the Earth has much in common with a well-constructed detective story. There is no shortage of clues pointing to the way in which the crime, the contamination of the pristine environment of early Earth, was committed. On the contrary, there are far too many clues and far too many suspects. It should be hard to find two investigators who agree on even the broad outline of the events that occurred so long ago and made possible the subsequent evolution of life in all its variety. [1]

On the other hand, the super-detective Sherlock Holmes, a character invented by the British author Sir Arthur Conan Doyle, while speaking to Dr. Watson, made the following declaration:

> When you have eliminated the impossible, whatever remains, however improbable, must be the truth. [2]

This purely logical principle of Sherlock Holmes equally applies to the emergence of life on young Earth, a historical process like a crime scene. We do not know what happened

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_1

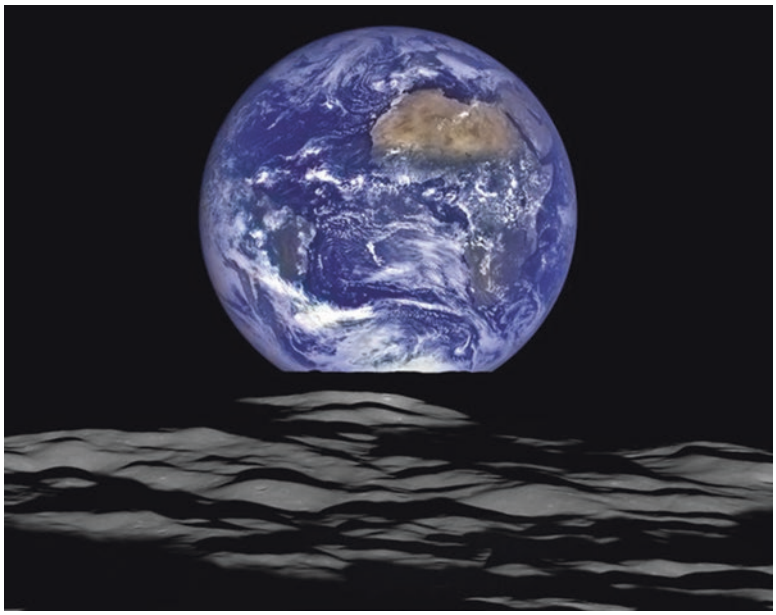


Fig. 1.1 Earthrise from the lunar surface against the blackness of space, captured on October 12, 2015, by NASA's 'Lunar Reconnaissance Orbiter' spacecraft; a stunning view that evokes the famous 'Blue Marble' image captured by astronaut Harrison Schmidt from Apollo 17, which also featured Africa prominently. (Courtesy of NASA)

four billion years ago. There were no witnesses, but all sorts of wild speculations have arisen. The life teeming on Earth today must have evolved from a simple common ancestor about four billion years ago, and yet there is no consensus on how life arose or even on what life is.

Before I begin my narrative, I want to acknowledge the many pioneers and trailblazers who have previously attempted to solve the mystery of life's origin. We must remember that earlier philosophers and scientists did not have the same knowledge that we now have owing to the recent developments in astronomy, geology, chemistry, genetics, and molecular biology. Here, I review some of the well-known theories of the origin of life from a historical perspective, acknowledging the insights and scientific reasonings of these prior thinkers. Until fairly recently, the origin of life had been a sacrosanct territory, the exclusive province of theologians.

1.2 Early Theories of the Origin of Life

People have pondered for ages how life originated on Earth. Theories abound, ranging from religious and scientific doctrines to other notions that border science fiction. Many philosophers and scientists, from the ancient Greeks to the eighteenth-century enlightenment thinkers, have suggested natural genesis ideas. The early Greeks proposed that living organisms originated by natural processes in sea slime through the action of heat, the Sun, and air. Revolutionary for its time and reminiscent of modern biogenesis, this secular concept was typical of rational-minded Ionian insight. Two centuries later, Aristotle proclaimed that certain forms of life generated spontaneously from decomposing matter and soil. Although rich in poetic metaphor, the first chapter of Genesis hardly puts forth the question of origin since, according to the Biblical account, God creates everything by divine fiat. Throughout the Early and Middle Ages of the Christian era, Aristotle's view on the genesis of life, with a shade of divine touch, was generally accepted by the learned. The idea that maggots in rotting meat, fleas in dung, and intestinal worms spontaneously arise from decaying matter did not belie the Holy Scripture. According to the 'spontaneous generation' theory, nature itself produces new living entities and living forms are continually being created from nonliving elements. In the seventeenth century, however, the invention of the microscope, instead of shedding light on spontaneous generation, served to darken it. This new tool opened a fantastic microscopic world never seen before. The microscope showed myriads of microbes, called 'animalcules,' in water, soil, and decaying substances. They seemed to appear from nowhere. The idea of spontaneous generation prevailed for a couple of centuries until Louis Pasteur dispelled it in 1862, demonstrating by meticulous experiments that bacteria are not formed spontaneously in a sterile environment. Instead, they form by division of preexisting cells. According to Pasteur's research, only life begets life; it cannot arise from nonlife [3].

However, Pasteur's conclusion also challenged Darwin's theory of evolution. If life does not arise from nonlife, then how did it emerge in the first place? Primarily concerned with the evolution of complex life forms from simpler ones, Darwin said little about the ultimate origin of life. Such ancient origins, he believed, were lost in the mists of time. His evolutionary theory was not about the beginnings of life

but about the processes that gave rise to biodiversity, i.e., the tree of life. However, in a letter in 1871 to his young botanist friend Joseph Hooker, Darwin mused:

> It is often said that all the conditions for the first production of a living organism are now present, which could ever have been present. But if (and oh what a big if) we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, etc., present, that a protein compound was chemically formed ready to undergo still more complex changes, at the present day such matter would be instantly devoured or absorbed, which would not have been the case before living creatures were formed.

In this letter, Darwin made a bold statement about the origin of life: life could have formed from organic molecules on sterile, lifeless ancient Earth. Darwin had moved away from the long-standing assumption that life was a gift from God or the gods. In other words, he suggested the creation of life from nonlife. Life could have arisen through geochemical processes. There was nothing in the prebiotic world that might have consumed or biodegraded these organic molecules, Darwin argued. These molecules could have survived and accumulated for an extremely long time to form the first cell. For the concentration of these biomolecules, Darwin intuitively suggested a small, warm, land-borne environment instead of a global salty ocean. In these small ponds, simple prebiotic compounds could have undergone further reactions, thus producing more complex molecules. Once life arose, it would have devoured the biomolecules from which it had sprung, using up all the available building blocks of life so that a second spontaneous origin would have been impossible. Thus, life could have arisen from nonliving chemicals only once. Darwin's contemplations set the stage for the next century's study of the origin of life, prefiguring the concept of the primordial soup. He even listed its ingredients, such as ammonia and phosphoric salts, and likely energy sources, such as light, heat, or electricity. Darwin's 'warm little pond' became a catchphrase for an incubator for prebiotic synthesis. It was a simple beginning, but it would become the basis of the first widespread hypothesis of how life likely began in a prebiotic environment.

During the same period, the great biologist Thomas Henry Huxley (1825–1895)—known as 'Darwin's bulldog' for defending his theory of evolution—coined the term 'abiogenesis' for the prebiotic synthesis of the origin of life from nonliving chemical systems through a gradual process of increasing complexity. Huxley had in mind that the chemical reactions of life slowly emerged on early Earth over a long time. He knew that the mixture would have to be more complicated than Darwin's ammonia and phosphoric salts in a 'warm little pond,' but he did not work out the details. According to Huxley, abiogenesis preceded biogenesis—the emergence of life. Thus, early on, Darwin and Huxley recognized the importance of 'chemical evolution': the production of simple precursors that could be assembled to generate life by abiogenetic processes.

1.3 Panspermia

The idea that life originated elsewhere and was then seeded here has had a long and checkered history. An Ionian Greek named Anaxagoras (510–428 BC) believed that life arrived on Earth as seedlings that came through space from other worlds. Anaxagoras called his hypothesis 'panspermia,' meaning 'seeds everywhere.' In 1909, Svante Arrhenius, a Swedish chemist and Nobel laureate, revived and elaborated the panspermia theory that life might have come to Earth in the form of tiny germs or spores propelled through space by solar wind. According to this view, microbial life exists throughout the universe, distributed by space dust, asteroids, and comets. Life was seeded from space, traveling through the interstellar emptiness from a living world elsewhere. Earth was thus a foster mother. Life began somewhere in the universe and was transported ready-made to Earth. Life hitchhiked across the galaxies and solar systems on meteorites and in interstellar dust and flourished when conditions became suitable, such as on favorable planets. Famous scientists, including Lord Kelvin, Fred Hoyle, and Francis Crick, have supported the panspermia theory, arguing that it would otherwise have taken too long for life to have originated on Earth by trial and error.

Controversial reports of fossilized microbes in the Martian meteorite ALH84001 rekindled the panspermia theory. This famous meteorite was picked up in 1984 from the pristine blue ice near Allan Hills, Antarctica. In 1996, NASA scientists publicized the discovery of chemical and fossil remains of microscopic organisms found in ALH84001, which had lived on Mars 4.5 billion years ago. They interpreted the tiny, filamentous, rod-like structures with segmented bodies within the meteorite as fossilized bacteria whose width was about a hundredth of that of human hair [4]. The announcement was a worldwide sensation. Most scientists are skeptical about whether ALH84001 contains any evidence of early Martian life. They believe that this putative microfossil is too tiny even for a bacterium and that this could be a preservable artifact. Such setbacks remind researchers that bacterial life usually does not leave a unique trace.

Most scientists utterly reject the original panspermia hypothesis, declaring that microbes have never been found in space. Space is cold, uninhabitable, and hostile to life. This theory circumvents the ultimate origin of life and moves it outside this planet to a more hospitable scene of origination. However, still, life must have begun somewhere and, as such, genesis still needs an explanation. The panspermia hypothesis does not attempt to solve the mystery of life's origins and therefore falls outside the scope of this book. However, recently, there has been a revival of some aspects of the panspermia idea.

In 1992, Christopher Chyba and Carl Sagan proposed the modern version of the panspermia hypothesis. This states that the prebiotic building blocks of life on Earth came from interstellar space and were exogenously delivered by meteorites. Still, that life itself was an endogenous production that was first synthesized in the womb of our planet [5]. In this book, I have endorsed and elaborated this view of Chyba and Sagan that primitive Earth may have received the necessary organic molecules via exogenous bombardment by asteroids, comets, and their fragments. These ancient impacts then jump-started life on our planet through the endogenous process in a suitable cradle such as hydrothermal impact crater lakes.

1.4 Abiogenesis: The Oparin–Haldane Theory

Although the origin-of-life question remained unresolved, there was considerable progress after Darwin's time. Most investigators believed that life originated on Earth and not in outer space. Based on this premise, nearly a century ago, independently of each other, two scientists began to tackle Huxley's question of abiogenesis—how could life arise on primordial, prebiotic Earth. In the 1920s, the Russian biochemist Alexander Oparin [6] and the British evolutionary biologist J.B.S. Haldane [7] subjected the origin of life to its first modern, comprehensive analysis. Oparin began by painting a picture of a violent Earth at the time of its formation. The planetary crust was searing hot as rocky matter impacted it from space. Eventually, however, Earth cooled down enough for water vapor to condense into liquid water, and the first rains fell. Both Oparin and Haldane believed that abiogenic materials could form organic molecules—that is, molecules necessary for the chemistry of life—in the presence of an external energy source. At that time, they believed that Jupiter was a good model for how planets began, with a thick, reducing atmosphere leftover from the formation of the solar system. They perceived early Earth, once it cooled down, as warm and wet and suggested that simple organic compounds, like those associated with life, were synthesized by natural chemical processes in a reducing atmosphere rich in methane, ammonia, hydrogen, and water vapor but lacking free oxygen. Ultraviolet (UV) light, lightning, and other forms of energy acted on these gaseous mixtures to create organic molecules. The absence of oxygen in the primitive atmosphere was crucial because, otherwise, any organic compounds that formed would rapidly be oxidized to carbon oxide (CO_2) and water (H_2O) and would thus not be available to participate in prebiotic processes.

Haldane likened Earth to a vast chemical laboratory powered by solar energy and called it the primeval sea. He suggested that, in the absence of living organisms to feed on organic compounds, the sea would have reached the consistency of a hot dilute 'prebiotic soup.' He hypothesized that abiogenesis took place through a complex mechanism involving enzymes and viruses. For his part, Oparin proposed that the accumulation of these organic molecules in the sea turned into floating 'coacervate' drops—spherical aggregates of lipid molecules with membranes—that formed protocells. Such compartmentalization was essential for prebiotic synthesis. Oparin's work with coacervates confirmed that enzymes—fundamental for the biochemical reactions of metabolism—functioned more efficiently when contained in membrane-bound spheres than when free in aqueous solutions. The 'prebiotic soup' was formed when these coacervates—which eventually gave rise to the first cells by the self-assembly of small molecules into larger, more complex molecules—began to concentrate in the oceans.

Oparin, the biochemist, contended that the cell-like coacervates that were synthesized by early Earth engaged in primitive, prebiotic metabolism. In contrast, Haldane, the geneticist, advanced the prebiotic synthesis of gene- or virus-like molecules. Oparin argued that proteins, contained in oily droplets or membranes, appeared first, followed by nucleic acids, and Haldane suggested that during prebiotic synthesis, nucleic acids were formed first and that proteins and membranes came later. These two views of abiogenesis—proteins first vs. genes first—are still contested today. The Oparin–Haldane models of abiogenesis explored the possibility that life formed spontaneously and naturally on early Earth. Their views were compelling, but there was no experimental evidence to back them up. Rather, the speculative work of Oparin and Haldane laid the foundation for other scientists to conduct laboratory experiments to test these hypotheses. Lab-based confirmation of their ideas would arrive a quarter-century later.

1.5 The Miller–Urey Experiment: Life in a Test Tube

The year 1953 may be called the 'annus mirabilis'—the 'year of wonders'—in the history of molecular biology and the origin-of-life research. It was in this year that James Watson and Francis Crick unraveled the double helix structure of a DNA molecule, Frederick Sanger determined the amino acid sequence of proteins, and Jonas Salk announced his successfully tested polio vaccine. In the same year, Stanley Miller and Harold Urey, at the University of Chicago, experimentally demonstrated that life's essential ingredients—amino acids—could be synthesized by simulating the primordial conditions of young Earth. Stanley Miller, then a graduate student, carried out a landmark experiment under the supervision of his famous Nobel laureate mentor, Harold Urey [8]. To simulate Earth's primitive atmosphere, Miller

placed a combination of methane, ammonia, hydrogen, and water vapor, as previously proposed by Oparin and Haldane, in a sealed glass apparatus. The energy input to the mixture was supplied by spark discharge and simulating lightning, with the entire mix circulating through a cooling tube, causing the gaseous combination to condense in imitation of rainfall over a pool of water. To synthesize the 'prebiotic soup' and mimic evaporation in the ocean, heat was supplied to the contained liquid. After a week, the experimental results were astounding. Three amino acids (monomers of proteins), including alanine, glycine, and aspartic acid, were spontaneously synthesized from inorganic raw materials. The undeniable conclusion was that the molecular components of life can indeed self-organize from simple precursors without a conscious direction. The Miller–Urey experiment—the first experimental breakthrough in the origin-of-life research—immediately received wide recognition and publicity. It confirmed that the chemical evolution of the critical ingredients of life on primitive Earth, under the kinds of conditions envisaged by Oparin and Haldane, could be simulated in a laboratory.

Moreover, preserved vials from the Miller–Urey experiment offered new hints on the origin of life more than a half-century later. About a decade before Miller died in 2007, his students recovered a set of boxes from his archives containing hundreds of vials of dried residues collected from electrical discharge experiments conducted in 1953 and 1954. These vials revealed 40 different amino acids and amines, far more than what Miller had initially reported in 1953 [8], demonstrating the potentially robust formation of important biological compounds under possible geochemical conditions [9].

The laboratory synthesis of amino acids, the monomers of proteins, created optimism among scientists that life could be created in a test tube. The Miller–Urey experiment inspired the birth of a new scientific discipline, i.e., prebiotic chemistry, and raised expectations that scientists could one day unravel life's origins with simple chemistry experiments. Many variations of the original experiment followed, as many enthusiastic chemists repeated and modified Miller-type experiments. A wide variety of the building blocks of life—amino acids, sugars, and RNA nucleobases—were created in laboratories. Using an assortment of recipes of the early atmosphere and different kinds of energy sources such as ultraviolet (UV) radiation and high-power shock wave, cell membrane components were synthesized. One of the most crucial experiments, by Joan Oró at the University of Houston, demonstrated the synthesis of adenine, a major component of RNA and DNA under prebiotic conditions [10]. However, although laboratory experiments in prebiotic chemistry have created numerous building blocks of life, including sugars, the purine and pyrimidine bases of nucleotides, and even lipid membranes, these experiments, fascinating as they are, still do not tell us how life began on early Earth.

Excitement over the Miller–Urey experiments has long since subsided. Importantly, however, the experiments resulted in a racemic mixture that contained both left- and right-handed amino acids. The real challenges now lie in selecting and concentrating left-handed amino acids and then linking these molecules to form chain-like protein molecules. The amino acids in these experiments never grew into more complex proteins. In recent years, it has become clear that life is more complicated than anyone had previously believed. Living cells, it turns out, are not just assemblages of polymers and macromolecules. They are vastly intricate, minutely integrated molecular networks with embedded biological information systems, far more stable and sophisticated than any human-made computer to date and beyond the reach of any analogy to the software/hardware dichotomy in computational devices (Fig. 1.2).

After 50 years of experiments in numerous laboratories around the world, few breakthroughs have followed these pioneering efforts. Scientists have produced the building blocks of life; they have not produced life itself. The recipe for brewing life is still missing. Some speculate that the missing ingredient is time. It took hundreds of millions of years of trial and error on early Earth to produce the first microbes from biomolecules, so how can we condense this long geological duration into short time in a laboratory? Processes that are impossible in a hundred days in a laboratory may yet be inevitable after a hundred million years of natural experiments. This is the order of time in which mountains rise, continents drift, and oceans form. Geological time is difficult, if not impossible, to simulate on the laboratory scale.

Moreover, many scientists suspect that the early atmosphere and prebiotic Earth were quite different from what Chicago's team supposed in 1953. Perhaps Earth was never teeming with hydrogen-rich gases like methane, hydrogen sulfide, or hydrogen itself. Instead, the primitive atmosphere may have consisted of stable mixtures of gases such as carbon dioxide and nitrogen rather than methane and ammonia [11]. One way to avoid the problem of the chemical state of the early atmosphere for the synthesis of organic materials is to forego an atmosphere entirely. An alternative energy source to surface-based systems is provided by geothermal energy at hydrothermal vents.

The discovery of amino acids and nucleobases and a wide range of organic compounds in meteorites suggests that the building blocks of life might have come from space, eliminating the need for chemical processes to produce them on Earth. The far more significant challenge, then, is to create nucleic acids—the building blocks of molecules like RNA and DNA—in the right environment. The origin of life, we believe, lies in the origin of these replicator molecules that

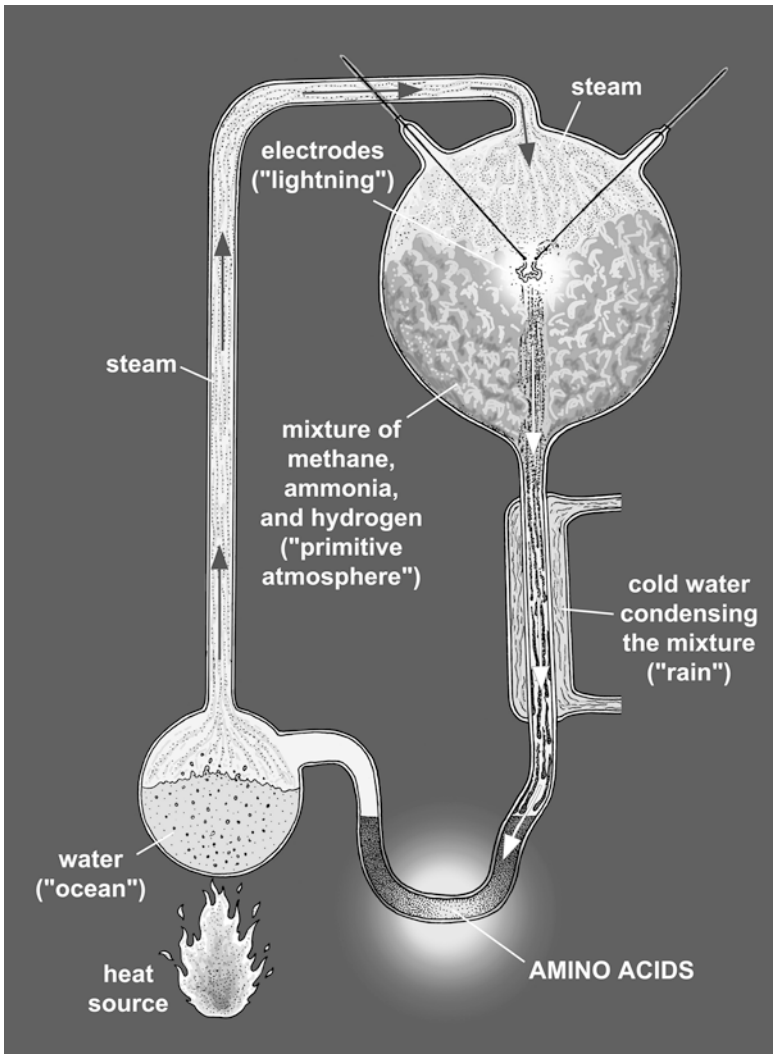


Fig. 1.2 The Miller–Urey simulation of prebiotic synthesis. The mixture of gases (methane, hydrogen, and ammonia) is circulated through the apparatus together with water vapor by boiling water in a flask. This mixture is subjected to an electrical discharge in a second condensing flask with a cooling jacket. The trap contains a variety of amino acids found in proteins

can make copies of themselves. If early Earth lacked some critical chemicals essential to forming life, did the building blocks of life come from elsewhere and then travel to Earth for blossoming?

1.6 Clue from Outer Space

Astrobiologists began to look for the building blocks of life in interstellar space and meteorites. One of the puzzling questions associated with life's origin is related to the source of these chemical building blocks of life. If they were not synthesized on young Earth, then perhaps the building blocks of life came from outer space during periods of heavy bombardment. Many of these complex biomolecules, including lipid membranes, amino acids, nucleobases, phosphorus compounds, and sugars, have been detected in carbon-bearing meteorites. Decades of study suggest that interstellar molecular clouds and circumstellar envelopes are factories of organic molecular synthesis that also occurred in the solar system ahead of life's origin on Earth [12]. These prebiotic organic compounds were delivered to early Earth by impacting comets or asteroids that contained the building blocks of life. With the exploration of space, the study of the origin of life has taken a broader perspective that includes planetary beginnings and astrobiological research.

As we develop our story in Chap. 4, we will learn that we, along with the world we inhabit, are recomposed from remnants of the massive explosions of spent stars in the Milky Way. Hydrogen in DNA itself has a cosmic connection to a universe three times older than our solar system. All life interconnects the cycles of Earth, the orbits of the planets, and the evolution of the stars. 'We are stardust, billion-year-old carbon,' Joni Mitchell famously sang in her song about the 1969 Woodstock festival. Recent astrobiological evidence has supported her poetic vision. We are truly stardust. This incredible song reminds us of our past heritage. 'And we've got to get ourselves back to the garden.'

1.7 Conclusions

Today, life is a dominant geological force with incredible biodiversity, but, when the planet formed, it was a sterile molten globe. How did lifeless Earth become a living blue planet? It is one of the most challenging questions of all—how did life on Earth begin four billion years ago? In 1871, Darwin discussed the issue in a letter written to his friend Joseph Hooker. He mused that life could have originated from 'a warm little pond' with all sorts of chemicals, such as ammonia and phosphoric salts, and various forms of energy, such as light, heat, and electricity, available on primitive Earth that underwent complex changes. It was a sketchy idea, but it would inspire Alexander Oparin and J.B.S. Haldane, in the 1920s, to suggest the first, far-reaching theory of abiogenesis. Oparin and Haldane independently proposed that simple biomolecules could have been created on early Earth in a reducing environment. They proposed that the primitive atmosphere was rich in methane, ammonia, hydrogen, and water vapor, probably released by volcanic eruptions, but it lacked free oxygen. Lightning, UV radiation, and other forms of energy acted on these gaseous mixtures to generate organic molecules such as sugars, amino acids, nucleic acids, and carbohydrates. These organic molecules began to float on the ocean surface as spherical vesicles or coacervates, precursors to the first cells.

In 1953, Stanley Miller and Harold Urey further advanced the concept of life's origin with a groundbreaking experiment. They simulated the primitive atmosphere in a sealed glass apparatus to recreate the primordial soup. They combined methane, ammonia, hydrogen, and water vapor in a flask, simulating early Earth's atmosphere, and exposed it to a continuous electrical discharge. Within a week, many amino acids commonly found in proteins, such as alanine and glycine, were produced in the flask. Miller's landmark experiment influenced other researchers in the origin of life to conduct many variations of his experiment for the next 60 years in numerous laboratories worldwide, without any breakthroughs. It was later realized that a Miller-type experiment might not be the right biochemical pathway for simulating the conditions of early Earth to analyze the origin of life.

Perhaps the essential building blocks of life came from interstellar space. Many of these complex biomolecules, including lipid membranes, amino acids, nucleobases, phosphorous, and sugars, have been detected in meteorites. These building blocks of life were delivered to young Earth by carbonaceous asteroids and comets during the Late Heavy Bombardment period. Life's origin on Earth, therefore, may have involved a combination of two distinct environments. One was the exogenous environment that comprised of the interstellar medium containing dust particles and meteorites. The other was the endogenous setting that ranged from ocean floors and submarine hydrothermal vents to the exposed continental landmasses with their terrestrial hydrothermal craters and hot springs. In the immediately following chapters, we look further into how asteroidal impacts may have ignited life on young Earth four billion years ago.

References

1. Orgel LE (1988) The origin of life—a review of facts and speculations. Trends Biochem Sci 33:491–495
2. Doyle AC (1890) The sign of four. Spencer Blackett, London
3. Oró JS (2002) Historical understanding of life's beginnings. In: Schopf W (ed) Life's origin. University of California Press, Berkeley, pp 7–45
4. McKay DS, Gibson EK, Thomas-Keptra KL, Vali H, Romanek CS, Clernett SJ, Chiller XDF, Maechling CR, Zare RN (1996) Search for past life on Mars: possible relic biogenic activity in Martian meteorite ALH84001. Science 273:924–930
5. Chyba C, Sagan C (1992) Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origin of life. Nature 355:125–132
6. Oparin AI (1924) Origin of life (translated from Russian). Dover, New York
7. Haldane JBS (1929) The origin of life. Rationalist Annual:148–169
8. Miller SL (1953) A production of amino acids under possible primitive earth conditions. Science 117:528–529
9. Bada JL (2013) New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. Chem Soc Rev 42:2186–2196h
10. Oró JS (1961) Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive Earth conditions. Nature 191:1193–1194
11. Kasting JF (2014) Atmospheric composition of Hadean-early Archean Earth: the importance of CO. Geol Soc Sp Paper 504:19–28
12. Pizzarello S, Shock E (2017) Carbonaceous chondrite meteorites: the chronicle of a potential evolutionary path between stars and life. Origins Life Evol Bio 47:249–260

Defining Life

2

'Music,' once said Isaac Stern, 'can be described, but not defined.' Perhaps the same is true of life itself.
—Antonio Lazcano, 1994

2.1 What Is Life?

The unique feature of Earth is the existence of life, and one of life's most beautiful mysteries is its diversity—the astonishing variety of life on Earth. Organisms occupy nearly every inch of the planet, from the scalding hydrothermal vents on the ocean floors to Antarctica's icy reaches. Such a variety of forms and habitats are a result of four billion years of evolution. We do not know how many kinds of organisms are inhabiting Earth right now. However, about nine million species of microbes, plants, animals, and fungi have been identified and named so far, ranging in size from microscopic bacteria to towering redwoods and colossal whales, each one occupying a particular niche in the web of life. It is estimated that we know only 7% of the species on Earth; the other 93% are yet to be discovered. Life permeates the biosphere, a thin membrane of organisms wrapped around Earth, about 21 km thick, stretching from the deepest point of the Mariana Trench extending down to 11,000 m below sea level to altitudes higher than the peak of Mount Everest, about 10,000 m above sea level. Microbial life is almost unbelievably resilient, with varieties that make a living in boiling springs, corrosive acid ponds, alkaline environments, salty marshes, deep rocks, deserts, stratospheric clouds, and oceanic vents. For all that, life's most significant event was its own birth.

However, what is 'life?' Although philosophers have been wrestling with the definition of life for millennia, modern astrobiologists agree that there is no universal answer. No matter what characteristic we specify to separate living from nonliving matter on Earth, we can find an example that contests this distinction. Does life reproduce itself? So does fire. Does life reproduce new forms? So do certain crystals in watery solutions. Nevertheless, before we probe more deeply into the origin of life, we should try to agree, as much as possible, on what life is. This seemingly simple but deceptively complex question leads to heated philosophical and scientific arguments. The answer is further complicated because researchers from different fields have differing opinions on what they believe should be included in a definition.

We have always been intuitively aware of the unity of life. Attempts to define 'life' are an ancient pursuit, going back to at least the time of ancient Greek philosophers. In 350 BC, Aristotle, in his treatise *De Anima*, offered what may be the first rigorous definition of life. In his account, life grows and maintains itself (he called this 'nutrition') and reproduces. He was correct, in that he emphasized two aspects of life—metabolic and informational—that we continue to consider as fundamental. In the seventeenth century, however, Rene Descartes argued that living things are to be understood as machines, a description that continues to be hotly debated. Today, although we seek to understand the living world in the context of genetics and evolution, a complete definition of life remains an enigma. We think we know a living thing when we see one, but it is still challenging to define it unambiguously.

2.2 Attributes of Life

Cells are the smallest unit of life. Many microbes consist of a single cell, about 1 micrometer in diameter. Self-evidently, such a minuscular system has all that is needed for life to carry out its operations. All living things share basic processes such as growth and reproduction. If something is alive, then it can grow and, during a certain phase of its existence, make a copy of itself. Living entities possess systemic organizations, coded into genes, capable of metabolism, reproduction, and evolution. What is required of living organisms for these three processes of metabolism, reproduction, and evolution to work? To begin with, there must be a structure able to stabilize the series of necessary processes. These processes must be supplied with the energy essential for their operations and the substances required for maintenance and duplication.

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_2

In metabolism, there is an incessant flow of energy and matter through a network of chemical reactions, enabling a living organism to regenerate, repair, and perpetuate itself continuously.

Moreover, concerning their environments, living systems are open systems. They regularly take food in and pass wastes back. Thermodynamically considered, they operate far from equilibrium, in that their processes maintain an energy differential that preserves their organization about relatively disordered environments. At the same time, living systems are organizationally closed, in that the membranes produced by their operations enclose and bind their metabolic and genetic networks. Physical sequestration from the environment is a prerequisite for living things. Under these stringent conditions, genes 'encode' a copy of instructions for self-assembly and then 'read' them to produce a daughter organism.

Life is common to all living beings. However, what is the commonality of life? As we know, living beings are made from cells, which are chemical factories, using sophisticated information systems. The Nobel laureate and biochemist Christian De Duve [1] summarized the major processes of life as follows: 'the ability of a system to maintain itself in a state far from equilibrium, grow and multiply, with a continual flux of energy and matter supplied by the environment.' All living things use environmental factors for self-support and development. This definition contributes to two distinct pictures of living organisms. One depicts a population of entities that, possessing a hereditary mechanism, survives and evolves by adapting to its environment. The other depicts a system whose complex structure is maintained by the energy and matter flowing through it. All life on Earth functions through three kinds of macromolecules: nucleic acids, proteins, and plasma membranes. Nevertheless, despite the complexity of life compared with the relative simplicity of inanimate objects, the distinction remains blurry at the edges.

For instance, viruses are distinctly biological, but they are not precisely alive. Rather, they constitute an indistinct threshold that has fascinated microbiologists for decades. Viruses have a genome and a protein capsid. Some are even enveloped in a lipid membrane. Just like living organisms, they possess the genetic code, and they evolve. However, viruses are excluded from living entities because they lack true cell membranes as well as any form of energy or metabolism. They lack ribosomes, the protein-making machine. Crucially, they cannot reproduce on their own. Instead, to make new viruses, they must infect living cells and hijack their translation machinery. Somewhere in Earth's history, a collection of hitherto inanimate molecules definitively crossed the threshold between living and nonliving and became animate organisms. Viruses inhabit a twilight zone touching on both realms. The problem of viruses underscores the wider lack of consensus regarding the definition of life. Nevertheless, we must have some working definition of life if we are to distinguish and identify the first living cells from their protocellular ancestors.

2.3 Attempts to Define Life

In 1944, before discovering DNA, the physicist and Nobel laureate Erwin Schrödinger defined life as that which 'avoids the decay into equilibrium' [2]. This definition refers to the second law of thermodynamics, which says that in a closed system, entropy—a measure of the dissipation of energy—always increases. However, living things, said Schrödinger, can postpone this cosmic trend. Thermodynamically, unlike the closed system of the physical universe, living entities are open systems that draw matter and energy from their surrounding environment and gain energy by metabolizing these nutrients. Life makes 'order from disorder.' However, the thermodynamic concepts fail to clearly distinguish between living and nonliving systems, complicating their use for demarcating life from nonlife. For instance, according to Schrödinger's definition, because crystals also take in energy and create order when they form elaborate lattices of particles, they should count as life. However, they can hardly be considered 'living' [3].

Schrödinger argued that organisms must produce and adopt modes of an informational order, an insight that we now associate with the genetic code. He speculated that genes constituted some kind of 'aperiodic solid' containing some sort of 'elaborate code script' that specified all the future developments of the organism. Following this argument, the mathematician John von Neumann noted a striking similarity between the biological gene and the metabolic apparatus, on the one hand, and the software/hardware distinction applied to the architecture of computer technology, on the other. He coined the phrase 'universal constructor' to name the computational simulation of the process of reproduction through cell division [5]. Today, we recognize that all living cells utilize DNA as the replicable repository of genetic information. They express this information by transcribing the DNA to the messenger RNA (mRNA) and then translating the mRNA into proteins and their cellular assemblages. Thus, the genetic code was the 'elaborate code script' conjectured by Schrödinger.

Living things propagate by producing approximate copies of themselves, thus creating new generations. In the process, life builds packets of order and complexity at the cost of expending large amounts of energy. In Schrödinger's phrase, living creatures 'drink orderliness' from their environment. The key to both features is genetic information: DNA molecules store the information needed to produce offspring. In developing his mathematical theory of communication, the

engineer Claude Shannon also approached the intimate link to energy in the inverse relation between information and entropy [4]. In both a living system and a communication system, more information means less entropy and higher organizational orders. Schrödinger's insights also inspired James Watson and Francis Crick, the molecular biologists who discovered the double helix structure of DNA. Watson and Crick embraced the idea that genes encode the information by which cells produce the proteins of which they are composed. Living systems store the genetic information that flows from DNA to RNA to proteins, eventually transmitting that genetic information to their offspring. The physicist Paul Davies has argued that genetic information, acting as cellular software, provides a singular definition of what life is [6]. It is the unique biological mode of information processing that separates life from nonlife.

In this view, the biological world is computational at its core, and life emerged at that instant when genetic information gained control over the first biomolecules. Instruction sets for the synthesis of proteins, analogous to computer algorithms, are found in every cell. Storage of the molecular information system is the key to defining life and understanding its origin. Yet, the algorithms guiding living processes constitute a system that is much more complex than today's most sophisticated computers. The embodied information system that emerged during prebiotic synthesis is still working perfectly four billion years later. Universally present in all life forms, information-directed protein synthesis may be life's unique signature from bacteria to humans. In the ongoing debates about what finally distinguishes life from nonlife, there is general agreement that life's informational aspect is a crucial property, perhaps the key property [7]. Life may be operationally defined as a system that both stores and processes the information necessary for its own reproduction. Following this definition, a virus would not be regarded as alive since it can only store but cannot process information. It must enter a living cell to do so. Within a living cell is an intricate information-processing machine. Life has a storehouse of knowledge in the form of language, code, and information systems embedded directly into the chemistry of the structures of its biomolecules that communicate using recursive, symbolic means.

Nevertheless, there are still many competing definitions of life. The mathematical physicist Freeman Dyson recognized that life requires a dual structure: a self-maintaining autocatalytic metabolic system that constitutes the hardware and genetic material that represents the software [8]. The Hungarian biologist Tibor Gánti argued along similar lines that a living system combines two subsystems, i.e., homeostatic and metabolic, to build and maintain its body. Second, it must have some sort of information storage system, such as genes, which could be copied and passed on to offspring [9]. The evolutionary biologists John Maynard Smith and Eros Szathmary describe life as 'any population of entities with the properties of multiplication, heredity, and variation' [10]. The information theorist Stuart Kauffman views life as either a single autonomous agent or a multi-agent system capable of self-reproduction [11]. Confronting these conflicting schemes, the Harvard biochemist and Nobel laureate Jack Szostak asserts that defining life is futile for understanding the origin of life [12]. He approaches the study of the birth of life as an effort to understand the transition from chemistry to biology. The many stages in the transition from nonlife to life can be instead seen to form a continuum within which defining a single point at which Darwinian evolution first emerged may not be possible.

Higgs [13] divides evolution of life into three stages, namely, chemical evolution, Darwinian evolution, and biological evolution. In chemical evolution, the diversity is generated by random chemical synthesis instead of (or in addition to) mutations, and selection acts on physicochemical properties, such as hydrolysis, photolysis, solubility, or surface binding. He argues that chemical evolution really is Darwinian evolution and not just chemistry but also establishes how chemical evolution differs from the usual kind of biological evolution that applies to genes and proteins in modern organisms. Darwinian evolution requires a mechanism for generation of diversity in a population and selective differences between individuals that influence reproduction. The ability to undergo biological evolution is considered a defining feature of life.

Kunnev [14] sharply distinguished chemistry from biology as a system that is capable of Darwinian evolution. He provides a clear definition of what is Darwinian evolution and what is life. The establishment of heredity in protocells in the peptide/RNA world defines the emergence of adaptation and diversification toward more complex structures, i.e., Darwinian evolution. He defines the minimum components required for Darwinian evolution such as the interaction between information and its corresponding structure/function during abiogenesis. Darwinian evolution has the innate ability to define transition from nonlife to life. The initiation of Darwinian evolution is the 'point of no return' after which life begins.

There are so many attributes of life to consider. As we have noted, all living systems have cell membranes, nucleic acids for replication, proteins for metabolism, and genes made of DNA. All encode proteins by way of specific chains of amino acids; all share a common energy currency through the synthesis of adenosine triphosphate (ATP). Life maintains itself not just by making more copies of itself but also by continuous self-production of the components out of which it assembles its structures. Many also consider this process of self-production, called 'autopoiesis,' to be a unique hallmark of living systems [15]. After all, even while DNA is a vital macromolecule for the informational activities we have been

discussing, that molecule itself is not alive. When a DNA molecule provides the template for the assembly of another DNA molecule, this is an event of structural replication. However, when a cell divides itself into identical daughter cells, this is the reproduction of entire systems. Incapable of reproduction, viruses are not alive; they are not autopoietic. Too small to self-maintain, viruses do not metabolize. They do nothing until they enter an autopoietic entity: a bacterial cell, a protist, a fungus, an animal, or a plant. For all their profound effects on the biosphere, viruses lack sufficient genes and translation functions to reproduce on their own. They persist as parasites or symbionts bound up with living, autopoietic systems. Here again, the definition of life is anything but straightforward. Life is not definable by singular units like genomes of cells but rather by a dynamic network of relationships among individuals, populations, and species. This network ultimately connects the entirety of our planet.

We have been reviewing the difficulties encountered in any attempt at a precise definition of life. Despite these difficulties, we need some definition of life if we are to identify its presence beyond Earth. Let us come back to a description of living functions. We have established that a living system integrates three critical interactive processes:

- The cell membrane maintains an identity over time by localizing all its components and protecting them from the environment.
- Cell metabolism uses free energy from its environment to maintain and augment the proteins of which cell components are constituted.
- Genetic information coded into DNA molecules carries inheritable information and controls cell division.

A cell must be able to copy and transmit all its genetic information to its daughter cells during reproduction. DNA and cytoplasm replications allow cells to do this. During cell division, however, occasional errors crop up in the replication of DNA molecules. The altered base sequences of these mutations then produce selectable variations in a population of cells, contributing to evolution by natural selection. The ability to evolve must be a critical aspect of any definition of life. In 1970, Carl Sagan suggested that 'life is a system capable of evolution by natural selection' [16]. Inspired by Sagan's idea, NASA proposed a 'working definition' of life as 'a self-sustained chemical system capable of undergoing Darwinian evolution' [17]. According to this definition, living entities display genetic variations—a key dynamic for Darwinian natural selection. This minimalist definition by NASA establishes a new level in the description of a biological system. Some researchers, such as Richard Dawkins, do not restrict Darwinian evolution to chemical systems but explicitly leave open the possibility that computers evolving through software mutations and variations may also constitute a form of life [18].

In 2011, the biophysicist Edward Trifonov analyzed the linguistic structure of 150 definitions of life, grouping similar words into categories and counting the ones used most frequently to produce a minimal or concise definition: 'Life is self-reproduction with variations' [19]. The 'variations' in Trifonov's definition are mutants, the result of mutations (errors in copying) that occur during reproduction, the raw material for Darwinian evolution. Although Trifonov's consensus and NASA's working definition do not use the same words, they share a central concept: 'life is capable of Darwinian evolution.' Moreover, NASA's definition has a universal application not limited to Earth, where DNA/RNA/protein-based life emerged. Although this definition could apply to a wide range of potential life forms anywhere in the universe, it also makes it hard to design a simple test for defining the very first life form or identifying a fossil.

Moreover, there are practical drawbacks to the Darwinian definition when searching for life on other planets: how long must we wait for a system to demonstrate that it is 'capable of Darwinian evolution' and under what conditions? Suppose we find what seems to be a microfossil in ancient hydrothermal chert in a crater basin on Mars. Can we use this Darwinian definition to ascertain whether this fossil is a product of primordial Martian life? Probably not. For that, we would also need some morphological attributes. Additionally, NASA's definition does not require the compartmentalization we see in all terrestrial cells. An exclusively informatic emphasis on genetic materials and their transcriptions—nucleic acids and proteins—overlooks the critical role that membrane structures play in life on Earth. Life is what is common to all living beings, and all living beings are made of cells; that is, they persist by means of their basic encapsulation apart from their environmental milieu. In prebiotic evolution leading to the origin of life, compartmentalization played the crucial role that it maintains even today. Thus, cell membranes and nucleic acids and proteins should be included as essential components of living systems.

Life is defined as an extension of being into the next generation and into the next species in evolutionary time. All life, from bacteria to humans, uses the same universal genetic code. The hereditary material of all known life on Earth is the DNA and the amino acids that constitute proteins. These biochemical similarities lead to the conclusion that life on Earth had a singular beginning, from which all subsequent life descended. Darwinian evolution explains how this common biochemical origin yielded such a fantastic diversity of living forms. If we ever discover extraterrestrials, then we have reason to expect that it will be based upon something like DNA, a self-replicating coding system, and it will have evolved along Darwinian lines. DNA is the key to the process of high-fidelity molecular replication and the viable variations subjected to Darwinian natural selection. In the

following definition of life, then, I have expanded NASA's version based on this expanded list of necessary attributes:

> Life is a self-sustaining, nucleic acid-based information system enclosed in a plasma membrane, which is capable of reproduction and Darwinian evolution.

We have based the definition of life on a single sample, i.e., Earth life. However, we need some answers based on this definition: Is it possible to have life without nucleic acid? I think possibly not. Is it possible to have life, which is not compartmentalized into a cell? I think possibly not.

In the case of the microfossil or extraterrestrial life mentioned earlier, if it retains a distinctive cell membrane and morphology, then the possibility of identifying primordial alien life becomes more robust.

There is no bounded entity smaller than a cell that is capable of independent reproduction. Thus, a hypothetical RNA cell, often depicted as the first cell, does not fit into our definition of life. We regard the RNA cell as a protocell, a long way from the first living cell. On its own, without the assistance of peptides, the RNA cell could not create the genetic code, the soul of the biological information system. Although replicating molecules, like nucleic acids, and catalytic molecules, like enzymes, are essential for life, they are not themselves alive. The properties of life arise only once these components become highly organized and sufficiently interconnected to gather specialized coordinated components into an interactive system.

We still do not understand what life is. Admittedly, our efforts to define life are based on only one example—life on Earth. Earth is the only planet in the universe known to possess life. Chris McKay of NASA suggests that the practical approach to searching for life beyond Earth is to determine what life needs [20]. The most straightforward list of environmental components needed to support life is probably carbon, liquid water, energy, and a few other elements such as nitrogen, sulfur, and phosphorus. De Duve claims that the constraints of physics and chemistry are so strong that life in any other place in the universe would be the same as that found on Earth. According to him, 'life is a cosmic imperative' [21]. Perhaps this universal view of abiogenesis makes a definition of life more feasible. Along these lines, the paleontologist Peter Ward coined the term 'Terroa' (derived from the Latin word *terra* for Earth) for Earth life [22]. We know what Earth life looks like from the single sample our planet provides, and, perhaps surprisingly, it is quite uniform. However, because we have only one example of life to go by, it is not surprising that we do not know for sure which features of Earth life are essential and which are just accidents of planetary history.

The philosophical question of the definition of life has increasing practical importance. As science is progressing toward understanding the origin of life on Earth, as laboratory experiments are approaching the synthesis of life, and as considerable attention is focused on astrobiology to detect life on Mars, on Jupiter's moon Europa, and on Saturn's moons Titan and Enceladus, the need for a general definition has grown apace. A definition of life is essential for understanding the origin and evolution of life on Earth and in the search for extraterrestrial life. If life is detected on Mars or other celestial bodies inside our solar system, then perhaps our definition of life will be refined, broadened, or modified. If physics and biology are tightly coupled, then life outside Earth may well be remarkably similar to life on Earth, at least at the microbial level. The biologist and biomolecular engineer David Deamer believes that an adequate definition of life will eventually emerge [23]. He does not consider efforts to define life as futile; instead, we are still adding parts to discover the whole. When scientists know enough to recreate life from scratch, they will know enough to define it

2.4 Conclusions

The definition of life has been a difficult problem to philosophers and scientists for millennia. From Aristotle to Carl Sagan, great minds have yet to come up with a definition that satisfies everyone. Since no unequivocal definition has yet been established, life is often defined by a list of attributes: life is composed of cells, has a life cycle, undergoes metabolism, can grow, adapt to its environment, can reproduce, and can evolve. Given this definition, because it cannot reproduce without a host cell, a virus is not a living entity. One prominent definition of life states that living systems are self-organizing and self-producing or autopoietic. From a physics perspective, living beings are open thermodynamic systems. Schrödinger suggested that life is an open system that exchanges matter and energy with its surrounding environment to make imperfect copies of itself. Life creates order from disorder. He also speculated that life must run something like a computer program, which is what we now call the genetic code. The key to both reproduction and the ability to create order in a living system is information. NASA has defined life as 'a self-sustaining chemical system capable of Darwinian evolution.' Although this definition of life avoids some of the simplest counterexamples that often defeat the 'list definitions' of life, it has some drawbacks. If we want to devise an experiment to find life, or if we find some microfossils in Archean sediments, then how do we observe Darwinian evolution in these cases? Our own view expands NASA's definition by adding the essential attribute of compartmentalization: Life is a self-sustaining, DNA-based information system enclosed in a plasma membrane, which is capable of reproduction and Darwinian evolution.

References

1. De Duve C (1991) Blueprint for a cell: the nature and origin of life. Neil Patterson Publishers, Burlington
2. Schrödinger E (1944) What is life? The physical aspects of living cell. Cambridge University Press, Cambridge
3. Catling DC (2013) Astrobiology: a very short history of introduction. Oxford University Press, Oxford
4. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423, 623–656
5. Von Neumann J (1966) Theory of self-reproducing automata. University of Illinois Press, Chicago. Edited and completed by A.W. Burks
6. Davies P (1999) The fifth miracle. Simon & Schuster, New York
7. Chatterjee S, Yadav S (2019) The origin of prebiotic information system in the peptide/RNA world: A simulation model of the evolution of translation and the genetic code. Life 9:25. https://doi.org/10.3390/life9010025
8. Dyson F (2004) Origins of life. Cambridge University Press, Cambridge
9. Gánti G (2003) The principles of life. Oxford University Press, New York
10. Maynard Smith J, Szathmary E (1999) The origins of life. Oxford University Press, New York
11. Kaufmann H (2000) Investigations. Oxford University Press, Oxford
12. Szostak JW (2012) Attempts to define life do not help to understand the origin of life. J Biomol Struct Dyn 29:259–266
13. Higgs PG (2017) Chemical evolution and evolutionary definition of life. J Mol Biol 84:225–235
14. Kunnev D (2020) Origin of life: The point of no return. Life 10:269. https://doi.org/10.3390/life10110269
15. Maturana HR, Valera FJ (1981) Autopoiesis and cognition: the realization of living, Boston studies in the philosophy of science, vol 42. D. Reidel Publishing, Boston
16. Sagan C (1970) Life. In: Encyclopedia Britannica, vol 22, 15th edn. Encyclopedia Britannica, Chicago, pp 985–1002
17. Joyce GF (1994) Forward. In: Deamer D, Fleischaker G (eds) Origins of life: the central concepts. Jones & Bartlett, Boston, pp xi–xii
18. Dawkins R (1983) Universal Darwinism. In: Bendal DS (ed) Evolution from molecules to men. Cambridge University Press, pp 403–425
19. Trifonov EN (2011) Vocabulary of definitions of life suggests a definition. J Biomol Struct Dyn 29:259–266
20. McKay CP (2014) Requirements and limits of life in the context of exoplanets. Proc. Natl. Acad. Sci. 111:12628–12633
21. De Duve C (1995) The beginnings of life on earth. Am Sci 83:428–437
22. Ward PD (2005) Life as we do not know. Viking, New York
23. Deamer D (2010) Special collection of essays: what is life? Introduction. Astrobiology 10:1001–1002

Biological Information Systems

3

Information is information, neither matter nor energy.

—N. Wiener, Cybernetics, 1965

3.1 The Concept of Information in Biology

All life shows entanglement between two highly different classes of polymers: nucleic acids (mRNA or DNA) and proteins. These two polymers can support each other because a highly specific and refined unidirectional information flow from genes to proteins is mediated by a code, the so-called genetic code. Nucleic acids store information in a linear sequence of nucleotides to make proteins. These unusual relationships between these two polymers show that information of genes and proteins is the specific linear order of their sequences. This process amounts to a transfer of information from genes to proteins. The discovery of the encoding of nucleic acids and their mode of translation into protein structures secured the modern view of biology as information science [1, 2]. Many of the everyday use of informational terms, such as transcription, translation, code, redundancy, synonymous, messenger, editing, proofreading, and library, have been incorporated into biology, conveying a similar meaning [3].

We use metaphors such as analog and digital, software and hardware, and nanobots and computers to compare life's information systems with human-made machines. This comparison is subtle if we understand the limitations of these metaphors. For example, digital information is like a 'program,' and an analog system is like a 'computer' that runs the life program. A cell is more complicated, reliable, and versatile than a supercomputer. The metaphors and analogies only explain a portion of the activities of the biological systems. As we trace the biochemical pathways for the origin of life, we see a continuum of different forms of information from the most basic and primitive to the most sophisticated in evolutionary terms [4].

3.2 Biological Information Systems

Biological information systems include all manifestations of information in living organisms. Life is an information-processing phenomenon at every level of organization, from molecules and cells to all living organisms. Any study of the origin of life must address the origin of biological information systems as well. Life can acquire, store, process, and use information to organize activities. Information is essential for life. It is one property, perhaps the key one, that separates life from nonlife. Information is the logic of life that makes a living system more organized, ordered, and complex. The way that information flows through and between biomolecules and cells is unique in nature [4]. Earth's biological and informational evolution are intertwined and inseparable. Life and its information systems form a closely coupled entity, influencing each other in a complex feedback loop that, through evolution, has acquired the ability to store and process information necessary for its growth, metabolism, and self-reproduction. Life transmits heritable information to its progeny and undergoes Darwinian evolution. Information does not change irrespective of whether it is encoded in nucleic acids or proteins: information is a substrate-independent concept. Information is central to a meaningful description of biological processes, but its status as a physical entity remains elusive. Darwinian evolution led to an increase in information content and a decrease in randomness during abiogenesis.

Traditionally, analog and digital biological information systems emerged about four billion years ago during abiogenesis [5–7]. De Duve [8] viewed the pathways of life as both determinate and directional, where the vector of evolution lies in the structural, informational, and catalytic molecules. It is well-known in biochemistry that biomolecules are

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_3

highly sensitive to changes in their environment—changes in pressure, temperature, pH concentration, adenosine triphosphate (ATP) concentration, molecular count, etc. These elements are in constant flux, enabling and forcing biomolecules to become a dynamic and flexible system to adapt and quickly respond to changes in the environment. Biomolecules have a reconfigurable internal structure that enables them to change their internal structure in the best way to face and solve the problem. They must also deal with limited resources and time. Analog computing is better suited for such a situation. It requires fewer parts, fewer resources, less energy, and less time than digital computing. Therefore, biomolecules—both large and small—are analog machines with their own embedded analog information system. They perform analog computing. It is highly instructive to understand the nature of a molecular analog information system. An analog information system's internal structure is not fixed like that of a digital information system. Instead, its internal structure is reconfigurable and solves the problem (situation) by changing its structure in a suitable way [9]. Contrary to popular belief, the major difference between analog and digital computing is not in their ability to process discrete versus continuous values but lies in their type of internal structure. The internal structure of a digital computing system is fixed. In contrast, the internal structure of an analog computing system is reconfigurable, and the analog computing system can change/reconfigure itself to deal with a given situation.

Each molecular unit has its own information and information system, meaning that 'information comes from within.' In other words, the molecular units may receive signals from the environment, but they contain their own information to process the signals. Each molecular unit performs its function using its own structure, information, and the information–function interdependency rules [10]. We assert that the information contained and used by various molecular units is in the form of four major categories—time, space, control, and energy. The time information consists of temporal information such as rate, clock, etc. The space information consists of spatial information such as pattern, proximity, attractiveness, sequence, etc. The control information includes signal and regulatory information. The energy information includes potential energy, charge differences, etc. The molecular units use, consume, and produce this information.

Molecular information systems began early in the interstellar medium in the building blocks of life, which were delivered to young Earth by meteorites during the Heavy Bombardment period. Analog information systems first appeared in abiogenesis, followed by digital information systems. These two information systems operating separately and in close cooperation streamlined the prebiotic synthesis from chaotic molecular assemblages and provided directionality to the flow of information [13].

A digital information system (DIS) includes genetic information that was slowly built by the coded sequences of nucleotides in mRNA in the peptide/RNA world. It is a latecomer in abiogenesis. Digital information is encoded in discrete linear sequences of nucleotides in mRNA and later in DNA. The sequence of nucleotides in mRNA and DNA determines the information content of the molecules—just as the sequence of letters in words determines the information content of words. Digital information processing in translation, the genetic code, and transcription are familiar to us. Less appreciated are the analog aspects of information processing.

The dichotomy between analog and digital information is not clear-cut. We show that between analog and digital, there is a transitional information stage, which we call the hybrid information system. The identification of these three systems helps us to document the coevolution of biomolecules and information systems during abiogenesis. These new approaches to prebiotic information systems are necessary for understanding the origin of life.

Living things collect and store information from their environment for survival. They adapt to their environment using the information to harvest energy and evade equilibrium. Life is characterized and sustained by several information-rich biological processes that govern cellular functions and significantly contribute to its overall complexity. Information is an important prerequisite for the onset of life. Any study on the origin of life must also address the origin of information systems. Prebiotic information would undoubtedly have been much simpler and was built incrementally over time.

Yockey [11] differentiated the processes of analog and digital information systems. Analog information is spontaneous, blended, and three-dimensional; components come from within the molecules. Contrariwise, digital information is linear, sequential, segregated, and is guided by coding rules. It is more robust and efficient to transmit information in digital form than in analog form. The digital code is inscribed in a template (mRNA or DNA) that provides the order in which the product (protein) is assembled. It is that order that specifies biological specificity. Linear, digital, and specific properties do not exist in the analog form. In digital information, both genes and proteins are manufactured or are artifacts produced by molecular machines, the former by the transcription machine and the latter by the translation machine [12]. A bilingual translation machine—a ribosome—orchestrates the translation of mRNA language to protein language. Life depends upon the interplay between both digital and analog coding, known as code duality. The analog reaction, on the other hand, is entirely in monolingual chemical language. The demarcation between nonlife and life is that the former is made of spontaneous objects, whereas the latter is made of manufactured objects or artifacts. Life is information plus meaning.

All life on Earth is possible because of a discrete digital mechanism of preservation and replication. Some of the information is encoded in genes and passed on from generation to generation.

Any modern complex computer operating system in the real world is partly digital and partly analog, and any living organism is an even more tangled mixture of digital and analog components. The concepts of analog and digital are far too narrow to encompass the subtleties of living cells. Here, the analog vs. digital dichotomy is more of an analogy than a precise description. Yet, it provides working tools to investigate the origin of information systems in life. We have described a third possible form, which is neither completely analog (because it is based on a discrete rather than a continuous component) nor completely digital (because it is noncoding) but a hybrid information system (HIS) that bridges the gap between the analog information system (AIS) and the digital information system (DIS) [13].

Although analog and digital information systems are well-known in the literature, we identified a transitional form, the hybrid information system (HIS), composed of noncoding RNAs. The HIS gave rise to major components of the translation machine, such as tRNAs (transfer RNAs), aaRSs (aminoacyl-tRNA synthetases), and ribosomes. These components originated from ribozymes [13]. HIS also contributed to the origin of viruses.

Walker [14] discussed 'biological memory' in the context of an information system as a mapping mechanism—a correspondence between the input and the resulting output of an event. In some cases, the input of the information carrier disappears automatically with the output of an event. For example, when translated into protein, mRNA is destroyed and recycled. DNA, on the other hand, is an information storage system, and it remains intact. It is generally regarded as a permanent memory mapping of protein structure. It relegates mRNA for protein building, where the recipe of a specific protein is laid down in a linear arrangement of the bases of transcripted mRNA or a gene. Most biological processes are one-way processes. The central dogma of information flow from DNA to mRNA to protein is no exception.

There are many examples of analog memory, such as self-assembly of bilayer membranes, folding of protein chains, or enzyme–substrate interaction. In the case of hybrid information, noncoding RNAs such as ribozymes and tRNAs play critical roles. Hybrid memory probably originated as a result of the Watson–Crick base pair rule that gave rise to RNA replication. Digital or genetic memory flourished during the buildup of mRNAs, translation, and the genetic code. It was permanently stabilized in the DNA/protein world with the establishment of the central dogma. We see the memory transfer during the reproduction of the first cells. Biological memory enables digital information to be stored, retrieved, and processed when needed. It may explain why the central dogma and the genetic code are universal in all organisms, from bacteria to human cells. Codon–amino acid mapping created a permanent 'memory bank,' which is analogous to the memory bank in the machine language translation system. This memory was permanently embedded in the last universal common ancestor (LUCA) four billion years ago and transferred to other domains of life during evolution. Molecular memory facilitates the processing of biological information.

Life depends not only on the flow of energy but also on the flow of information. A living system stores information and processes and uses that information to self-maintain and perpetuate itself. Life is an information-processing system in which memory is maintained in analog, hybrid, and digital forms. Information is the essence of life, but it is not synonymous with life. A living system must hold information and process and use it [15]. Often we interchange between the information processing molecule with its type of information system it contains. For example, we say DNA is digital. In reality, DNA is the storage or information processing machinery where the digital information is embedded.

Evolution is an information-generating and transmitting process. It creates information in a hierarchical structure and involves constraints, specializations, and symbiotic relationships. The origin of life has produced organic molecules of increasing size and complexity in collaboration with information systems through time. How can a living system emerge from a chaotic assemblage of space molecules in the hydrothermal vent environment? What rule might have guided the prebiotic synthesis? The rule of life is its information system. It reduces the number of possible prebiotic interactions in the hydrothermal vent environment and compresses the evolutionary goal of reproduction [16]. It provides directionality during abiogenesis.

Information is the currency of cellular activity and is the key to understanding the mystery of life's origin. Yet, its status as a physical entity remains puzzling. We have suggested that the molecular origin of life and its information systems are intimately linked in a complex feedback loop by hierarchical organization [13].

3.3 Where Did Life's Information Systems Begin?

Information has a physical basis through thermodynamics. Life is an open system and interacts with its environment—life exchanges matter, information systems, and energy with its environment for growth and nourishment. Living systems respond to the continuous environmental signals by complex computations they encounter. Life avoids decay via the second law of thermodynamics by importing information or negative entropy from its surroundings. It

grows by concentrating within itself and exporting entropy. The source of biological information, then, is the organism's environment. Both metabolism and reproduction are driven by information flow from the environment to the living system [17].

Recent evidence from carbonaceous asteroids has suggested that life's building blocks and water came from space during the Late Heavy Bombardment period (see Chap. 5) but were synthesized in the womb of our planet [18, 19]. Terrestrial hydrothermal impact crater lakes have been proposed as the prime habitat or the cradle of life [20] (see Chap. 6). A study of the Murchison meteorite suggests that these organic molecules developed analog information in the interplanetary medium [21]. These building blocks of life were delivered to Earth with an embedded analog information system (AIS). There was a transfer of the information site from the interplanetary medium to terrestrial hydrothermal crater lakes on young Earth during the onset of abiogenesis (Fig. 3.1) [13].

3.4 Destruction of Information: The Key to Abiogenesis

Life is a highly intelligent computing machine. There are two types of computing transformations, one in which information is preserved and one in which information is destroyed [17]. Examples of the former during biogenesis are the emergence of the genetic code, the first genes, translation machines, the first proteins, ancient viruses, DNA, and many more. However, many steps to achieve these prebiotic synthesis breakthroughs are entirely erased in the first cells. For example, aa-tRNA is not required to create a new gene or retrovirus for turning mRNA into DNA in the first cells. DNA creates mRNA directly by transcription. All these final products were transmitted vertically from the parent cells to the daughter cells. Too much raw data was generated during abiogenesis, which was continually destroyed, feeding the selective results to the next level. Similarly, mRNA instructions for making proteins are fleeting—they quickly self-destruct after being read by ribosomes. The value of computing is precisely in its ability to destroy information selectively, like house cleaning, preserving only those critical information-bearing algorithms required for survival. This is why the origin of life study is too difficult because of a large amount of missing data, which was destroyed during abiogenesis.

3.5 Transfer of Information

In abiogenesis, as we proceed, we will encounter the transfer of information from one molecule to another. There are two kinds of transfer of information: horizontal transfer of information (HTI) and vertical transfer of information (VTI). HTI is the lateral movement of information between two molecules, such as DNA to mRNA during transcription, mRNA to proteins during translation, tRNA to mRNA during encoding, virus to cell during infection, or environment to protocell during growth and nourishment. VTI, on the other hand, is the vertical transfer of information, such as from a parent to daughter cell. Transfer of information would play a critical role in prebiotic synthesis.

3.6 Prebiotic Information Systems

Prebiotic information systems are more complex, elegant, and advanced than human-made computing systems. There are two cycles of prebiotic information systems. In the beginning, prebiotic information systems were predominantly analog, and then biomolecules advanced to hybrid computing, which gave rise to the digital information stage in the peptide/RNA world [13]. The coevolution of biomolecules and prebiotic information systems is summarized in Fig. 3.1. The outcome of the first cycle of prebiotic system is the mRNA-directed protein synthesis from the cosmic ingredients. The flow of information is: AIS –>HIS–>DIS–>HIS–>AIS. The second cycle of information is summarized in Chapter 20. It began with mRNA protocells that gave rise to the first cells via viral infection (Fig. 20.2).

3.7 Conclusions

The molecular origin of life and its information systems are intimately linked in a complex feedback loop by hierarchical organization. The way that information is stored, processed, transmitted, and translated in biomolecules led to greater complexity in prebiotic synthesis. Biological information systems exist in three forms: analog, hybrid, and digital. These tripartite information systems were nourished from the chemicals and energy from the hydrothermal crater vent environments and permeated through the biomolecules. The analog information system (AIS) was manifested early in abiogenesis in the differing concentrations of chemicals. The AIS was embedded in the cosmic building blocks, imported from space, and proliferated in terrestrial hydrothermal crater vent environments in young Earth. Analog information created noncoding RNAs by polymerizing nucleotides that gave rise to the hybrid information system (HIS). The HIS was transitional and used components of both analog and digital information systems. It arose as an emergent property in the peptide/RNA world. The hybrid components would build the translation machinery components step by step. With the advent of mRNA molecules, the digital information system (DIS) became the dominant system. With the genetic memory residing in the digital sequences of mRNA, a

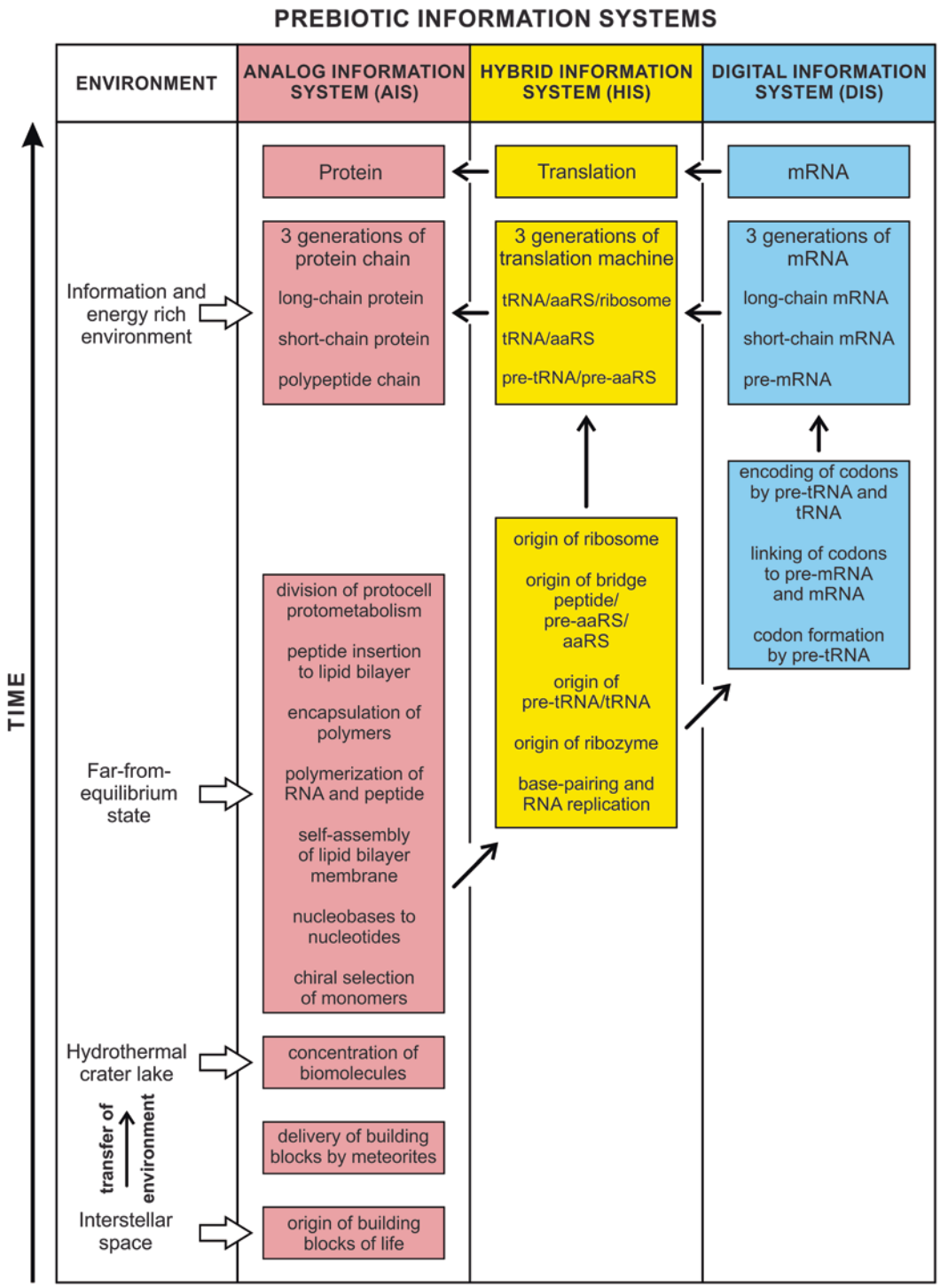


Fig. 3.1 Coevolution of biomolecules and prebiotic biological information systems. An analog information system dominated the early stage of abiogenesis. During the Late Heavy Bombardment period, there was a transfer of the site of information from interstellar space to hydrothermal crater lakes on young Earth. With the emergence of nucleotides, a hybrid information system began to emerge. The origin of pre-mRNA and mRNA marked the digital revolution. During the origin of translation and the genetic code, the directionality of information flow from mRNA to proteins emerged. (After Ref. [13])

mapping mechanism was developed between each codon and its cognate amino acid. As more and more codons 'remembered' their respective amino acids, this mapping mechanism developed a primitive form of the genetic code. The unidirectional information flow from mRNA to proteins produces analog output. Transfer of information, both horizontal and vertical, from one molecule to another is an important component of abiogenesis.

References

1. Küppers BO (1990) Information and the origin of life. MIT Press, Cambridge, MA
2. Von Neumann J (1966) Theory of self-reproducing automata. University of Illinois Press, Chicago. Edited and completed by A.W. Burks
3. Maynard Smith J (2000) The concept of information in biology. Philos Sci 67:177–194
4. Chatterjee S, Yadav S (2019) The origin of prebiotic information system in the peptide/RNA world: A simulation model of the evolution of translation and the genetic code. Life 9:25. https://doi.org/10.3390/life9010025
5. Walker IW, Davies PCW (2012) The algorithmic origin of life. J R Soc Interface. https://doi.org/10.1098/rsif.2012.0869
6. Koonin EV (2015) Why the central dogma: on the nature of the great biological exclusion principle. Biol Direct 10:52. https://doi.org/10.1186/s13062-015-0084-3
7. Baum DA, Lehman N (2017) Life's late digital revolution and why it matters for the study of the origins of life. Life 7:34. https://doi.org/10.3390/life7030034
8. De Duve C (2005) Singularities: Landmarks on the pathways of life. Cambridge University Press, New York
9. Ulmann B (2013) Analog computing. Oldenbourg Wissenschaftsverlag. https://ereader.perlego.com/1/book/654438/20
10. Kunnev D (2020) Origin of life: the point of no return. Life 10:25. https://doi.org/10.3390/life10110269
11. Yockey HP (2005) Information theory, evolution, and the origin of life. Cambridge University Press, Cambridge
12. Barbieri M (2016) What is information? Philos Trans R Soc 374:20150148. https://doi.org/10.1098/rsta.2015.0060
13. Chatterjee S, Yadav S (2022) The coevolution of biomolecules and prebiotic information systems in the origin of life: A visualization model for assembling the first gene. Life 12:834. https://doi.org/10.3390/life12060834
14. Walker I (1972) Biological memory. Acta Biotheor 21:203–235
15. Dyson F (2004) Origins of life. Cambridge University Press, Cambridge
16. Grisogono AM (2017) (How) did information emerge? In: Walker SM, PCW D, GFR E (eds) From matter to life: information and causality. Cambridge University Press, New York, pp 61–96
17. Davies P (1999) The fifth miracle. Simon & Schuster, New York
18. Pizzarello S, Shock E (2017) Carbonaceous chondrite meteorites: the chronicle of a potential evolutionary path between stars and life. Orig Life Evol Biosph 47:249–260
19. Bernstein MP, Sandford SA, Allamonda LJ (1999) Life's far flung raw material. Sci Am 263:42–49
20. Chatterjee S (2018) The hydrothermal impact crater lakes: the crucibles of life's origin. In: Kolb VM (ed) Handbook of astrobiology. CRC Press/Taylor & Francis, Boca Raton, pp 265–295
21. Deamer DW (2011) First life: discovering the connections between stars, cells, and how life began. University of California Press, Berkeley
22. Kurzweil R (1999) The age of spiritual machines. Penguin Books, New York

Origin of Life: A Model of Hierarchical Complexity

4

Evolution is hierarchical, operating not only on genes but on species.
—Stephen Jay Gould, The Structure of Evolutionary Theory, 2002

4.1 The Complexity of Living Systems

Life is an astonishing example of structural and functional complexity at work. The complexity of living systems occurs at every level in life's hierarchy, from numerous components to cells and organs to organisms and populations. Living beings develop through an incredibly complex series of interactions involving many different components, or subsystems, in the form of networks in an orderly and organized manner. These components are themselves made up of smaller molecular components, which independently exhibit their own dynamic behavior. Yet, when combined to form larger functional units, what results is utterly new and with new properties that emerge unpredictably. Like complexity itself, such emergence of new properties in living systems also occurs at every hierarchical level. This phenomenon—in which components join to form larger, stable structures having new properties that could not have been predicted from the characteristics of individual parts—is known as self-organization through spontaneous self-assembly processes. These processes are not under regulatory controls. The view of living systems as self-organizing networks of interconnected and interdependent components offers a compelling overview of the origin and evolution of life. Living systems are complex at every hierarchical level, from their metabolic capacities to bind or release energy and sensory capabilities to respond to external stimuli to organismal abilities to grow, differentiate their tissues, and self-reproduce. Many of these processes are mediated and fine-tuned by analog and digital information systems. Such overarching complexity may be the most characteristic feature of living systems. In hierarchical biological systems, one may even encounter downward causation, in which events at higher levels of a system may propagate effects to its lower-level components. No doubt, the origin of life from nonliving matter was itself the original expression of life's hierarchical complexity.

The origin of life on early Earth is one of humanity's significant unresolved scientific problems. Life's beginnings were the climax of a long history of prior cosmochemical and geological processes. Our ignorance about life's earliest period stems in large part from the fact that whatever events turned nonliving matter into the first living organisms occurred roughly four billion years ago and left no definite traces behind. We know that life emerged on our planet, but we do not know when, where, or how. There is no record of that time; there are no transitional fossils of abiogenesis. Life has left no clues at the scene of its birth. What we must do, then, is to work backward from contemporary microbial life, using scientific imagination guided by astrobiological insights. The oldest fossil evidence from the Archean hydrothermal sediments suggests that microbial life on Earth existed at least 3.8 billion years ago or even earlier [1]. This means that life must have begun around 500 million years after Earth emerged from the solar nebula. Life's advent marked the beginning of the Archean eon when the planet had cooled down enough for a global ocean to precipitate from the steamy primeval atmosphere [2].

The transition from inanimate matter to life, from simple organic compounds to living cells, was not a single event. Instead, it gradually increased in complexity, starting from cosmochemical events to new biological phenomena. By simulating the significant stages of abiogenesis, ranging from cosmic chemistry and chemical evolution to the synthesis of information systems and protocells, critical experiments conducted in controlled laboratory settings have narrowed the significant gaps in our understanding of the origin of life. Nevertheless, we still have a long way to go if we succeed in decoding the precise pathways leading to the origin of life. During the last few decades, experimental and theoretical progress has shed new light on the plausible ways of abiogenesis. Yet, the sheer

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_4

diversity of opinion in origin-of-life research suggests that we are not close to achieving a consensus on the matter. However, as we reason, experiment, observe afresh, and build up our multidisciplinary knowledge, our attempts to reconstruct the emergence of life are as worthy of pursuit as any other major scientific theory regarding the natural world. This is how science works: new data, new experiments, and new ideas challenge previous theories and older ideas. When a conflict arises between evidence and theory, a new, better theory results. The strength of a scientific theory rests on the simplicity with which it can explain the diversity of phenomena.

New evidence from astrobiology has led to a renaissance in origin-of-life research [3]. As a cosmological object, early Earth was not a closed system but received input from space and the surrounding environment. Both the organic and inorganic chemical components of early Earth originated from interstellar space. Carbonaceous chondrites, which date back 4.6 billion years, have been considered a clue to the early solar system and hence to early Earth. This book has borrowed a set of themes and ideas from astrobiology to propose a new way of understanding the origin of life.

The presence of life's building blocks in meteorites heightens our interest in them as carriers of the ingredients of life [4]. Via meteorite impacts, carbonaceous asteroids and comets could have shipped to Earth rich loads of elements and organic compounds crucial to life. The central theme of my narrative is that asteroidal impacts during the final phase of what is called the Heavy Bombardment, a geological period that occurred about 4.1–3.9 billion years ago, hastened the origin of life on young Earth [5, 6]. Importantly, meteorite impacts created innumerable hydrothermal crater lakes upon the Eoarchean crust. The young crust inadvertently became ideal cradles for concentrating exogenous organic molecules to form complex chemical systems. As the crater basins filled, volcanically driven geothermal vents heated the water and created convection. The resulting currents constantly stirred the water till it reached the consistency of a thick primordial soup, leading to the creation of the first simple organisms [7, 8]. Cataclysmic collisions from space may have set the stage for the rise of life on Earth. The environment of early Earth was undoubtedly not an Eden, but once the meteoritic onslaught was over, the conditions were right for life to take hold.

4.2 The Hierarchy of Prebiotic Synthesis of Organic Compounds

In the coming chapters, I shall retrace what I consider to be, in light of the present knowledge, the most plausible course for the natural development of life on Earth. New evidence from astrobiology provides a systematic exploration of how cosmic molecules in hydrothermal vent environments progressively self-assembled to create hierarchical structures with increasingly complex functions, all the way to reproducing living cells. Taking as my starting point the primordial delivery of the building blocks of life by asteroidal impacts on the Eoarchean crust about four billion years ago, I propose that life arose through five successive hierarchical stages of increasing molecular complexity—'cosmic,' 'geological,' 'chemical,' 'informational,' and 'biological' (Fig. 3.1). We model the origin of life as a sequence of emergent steps, each of which added elemental complexity to the geochemical world. Emergent complexity arises as energy flows through systems of numerous interacting agents and information systems evolve. As each stage continued to merge with the following stage, each node in the hierarchy of abiogenesis achieved an increase in organizational and evolutionary complexity. There was no blueprint for this biosynthesis; no foresight was exercised. Rather, each subsequent stage selected emergent modifications of what had come before in accordance with the universal laws of thermodynamics.

According to my reconstructions, the origin of life required at least five such steps:

1. The formation and delivery of cosmic biomolecules.
2. The presence of cradles for biosynthesis.
3. The chemical evolution of macromolecules.
4. The development of genetic informational systems through molecular encoding.
5. The emergence of the first cells capable of self-maintenance, reproduction, and Darwinian evolution.

Each primary stage involved molecular mechanisms that controlled selection from prior, lower-level organizations. Since reversions to lower-level behaviors can be deleterious to the complex entities attaining higher-level organizations, the formation of this multilayered hierarchy provided a kind of quality control at each assembly stage. The present-day clues for these primal hierarchies come from outer space, from our Moon and the inner planets of the solar system, from the geology of early Earth, from prebiotic chemical evolution, from laboratory experiments, and especially from the cellular components of contemporary life [4, 5].

4.3 The Cosmic Stage

In recent years, origin-of-life researchers have witnessed the most exciting advances in astrobiological development of cosmochemistry, i.e., the chemical analysis of extraterrestrial materials. Planetary materials are pieces of condensed matter left over from the time of the origin of the solar system, about 4.6 billion Ga. Such materials include interplanetary dust particles, pre-solar grains of matter, primitive meteorites, and regolith or stony matter from the Moon and asteroids.

Although life is most likely a phenomenon localized on Earth's surface, it is understandable only in its cosmic

connection. In short, life formed by itself from stardust not long, by cosmic standards, after Earth formed and cooled. Developing a full understanding of life on Earth appears to demand an understanding of its links to its cosmic environment. In the 'cosmic' stage, we can say that the building blocks of life started in the interstellar icy grains composed of gas and dust. The explosion of a nearby supernova, many times more massive than our Sun, formed these organic molecules as a byproduct [9]. The supernova expelled this material across space to form interstellar clouds, which we know to be factories of organic molecular synthesis. During the origin of the solar system, this interstellar organic material was incorporated into interstellar dust, comets, and carbonaceous asteroids. Meteorite impacts during the early history of our solar system seeded every planet with prebiotically significant organic compounds. As we have noted, the shipments of organic molecules delivered to the newly formed crust of young Earth during the Heavy Bombardment phase played a pivotal role in life's origin [10]. The surface of Earth was still sterile but became more complex with the meteoritic addition of organic compounds.

The meteorites that fall to Earth even today carry a message from space about the history of the solar system. Asteroids formed extremely early, approximately 4.6 billion years ago, and provide a record of the chemical processes in the solar system before life began. In a particular class of asteroids called carbonaceous chondrites, the element of carbon is abundant. Carbonaceous chondrites provided a sophisticated suite of organic compounds and water that were biologically relevant to the origin and sustenance of life. Meteorite impacts shipped these 'seeds' of life to young Earth. For instance, the Murchison meteorite, a carbonaceous chondrite that fell in Australia in 1969, has yielded many organic molecules, including sugars, phosphates, nucleobases, amino acids, and membrane-forming compounds [11, 12]. Comets, as well as meteorites, contain prodigious quantities of such organic components [10].

To sum up the importance of the cosmic stage, in this view, life arose from stardust, the priceless debris of a supernova explosion. The origin of life is a unique product of two worlds: the exogenous delivery of the building blocks of life by meteorites to early Earth, eventually leading to the endogenous production of the first life in the cradle of our planet [13]. The discovery in the 1960s and 1970s of a wide variety of organic molecules—amino acids, carboxylic acids, hydrocarbons, sugars, phosphorous, and purine and pyrimidine nucleobases—in carbon-rich meteorites, comets, and interplanetary dust particles supports the thesis of the extraterrestrial delivery of the building blocks of life rather than their endogenous production on primitive Earth itself [4, 10, 12]. The origin of life may have had interstellar beginnings during planetary formations, but meteorite impacts jump-started life by delivering these crucial biomolecules to Earth's surface. The interstellar dust and meteorites provided the building blocks of life, but our Goldilocks planet provided the ideal cradle for the biosynthesis that converted sterile molecules into living entities. From cosmology, geology was born. Next, in the geological stage, these abiotic, monomeric organic compounds accumulated to form suitable crucibles on early Earth, which were exposed to chemical energy that encouraged subsequent prebiotic synthesis.

4.4 The Geological Stage

By the beginning of the Archean eon, about four billion years ago, the global temperature had fallen to a point at which water vapor could condense into vast, shallow global oceans virtually covering the planetary surface, with scattered continental islands or protocontinents. Carbon dioxide and nitrogen dominated the atmosphere [2, 6]. Exactly where life first arose on this planet is one of the keys to the discovery of how life originated. Was there some sequestered basin where life's ingredients could mingle without being broken down? The answer may be found in the 'geological' stage. Most scientists now view hydrothermal vents as the most promising site for the origin of life, where superheated fluids rich in transition metal ions and hydrogen sulfide (H_2S) mix abruptly with cold water and precipitate out. The volcanic sediments surrounding the vents may include clay minerals, pyrites, sulfides, zeolites, carbonates, and amorphous volcanic glass, all of which are essential catalysts for prebiotic synthesis. Today, these vents support abundant life, much of which is independent of solar energy. Hydrothermal systems can develop anywhere on the crust where water coexists with a magmatic heat source. Hydrothermal sites are geochemically reactive habitats, far from chemical equilibrium, where hyperthermophiles thrive around superheated water supporting chemosynthetic (rather than photosynthetic) ecosystems. The chemicals found in these hydrothermal vents and the energy they provide could have fueled many of the prebiotic reactions necessary for the evolution of life.

Currently, there are two opposing hypotheses for the location of the primordial cradles of life: submarine hydrothermal vents [14–17] and terrestrial hydrothermal impact crater lakes [6–9, 18, 19]. Recent debates have raged over whether life began at sea or on land. In both submarine and land-based hydrothermal settings, mineral catalysts have facilitated reactions now performed by enzymes. In both locations, chemical reactions were far from an equilibrium state, with a source of energy available in the vent environment to drive synthetic reactions. A process to concentrate cosmic chemical ingredients sufficiently enough to undergo chemical and physical reactions is a prerequisite for prebiotic synthesis. This question of concentration of biomolecules may determine which early environment—marine vs. terrestrial hydrothermal vents—was the more likely incubator for life's beginnings [6]. The submarine hydrothermal vent hypothesis

has considerable appeal but lacks universal acceptance. Many aspects of the proposed scenarios remain questionable in the face of geological evidence of early Earth's environment and the chemical constituents of living cells. This theory suffers from the 'dilution problem' regarding precise organic compounds. In the vast Eoarchean ocean, the cosmochemical ingredients would be dispersed and diluted rather than necessarily concentrated, thus preventing them from assembling into the complex molecules necessary for life.

One crucial prerequisite for the origin of life is that simple organic molecules must be concentrated in the vent environment to becomes complex molecules. It is hard to see how, in an open ocean, cosmic and terrestrial chemicals could have mixed and concentrated, selected, and organized into even more complex molecules. Moreover, many macromolecules found in living cells, such as lipid membranes, RNA, proteins, and DNA, are polymers and form via condensation reactions that require an environment fluctuating between wet and dry. Dry states shift the thermodynamic equilibrium compared to reactions in water, and this alternation drives condensation–dehydration reactions. Wet and dry cycling occurs every day in land-based hydrothermal fields. This allows for the concentration of reactants and for polymerization. Submarine hydrothermal vent environments lack such wet and dry cycles [11].

Recent research has indicated that the chemistry of modern cells mirrors the original environment in which life first evolved. The chemical nature of terrestrial hydrothermal vents and hot springs resembles the composition of the cellular cytoplasm more closely than does the open ocean environment. By this logic, protocells must have evolved in habitats with high K^+/Na^+ ratios and relatively high concentrations of zinc, manganese, and phosphorous compounds, as in terrestrial environments. In contrast, seawater has 40 times more sodium than potassium, which could have inhibited protocell formation. The ionic composition leading to the origin of living cells could not have existed in marine settings but is somewhat compatible with inland geothermal systems, such as hydrothermal ponds and crater lakes [20].

Similarly, nucleobases distributed by meteorites and interstellar dust particles to early Earth could be polymerized into RNA only in terrestrial environments such as warm little ponds or crater lakes, providing dry and wet cycles, an operation that could not occur in the deep ocean [21]. Moreover, there are major tectonic problems with the submarine vent hypothesis. It is difficult to explain how submarine hydrothermal vents could offer a likely cradle for life on one-plate Eoarchean Earth. Without the activity of plate tectonics, how could submarine hydrothermal vents begin? Today, they are located along or near the axis of the spreading tectonic ridge. However, since plate tectonics only started to appear around 3–2.5 billion years ago [22], the Eoarchean oceans offered no spreading ridge for hydrothermal vent formation. In our view, to find the cradles of life, we will have to seek land-based hydrothermal systems.

Hydrothermal impact crater lakes on the crust of young, one-plate Earth appear to be the most likely environments for the beginnings of life [5–8, 18, 19]. During the late phase of the Heavy Bombardment period, impact events on the Eoarchean crust produced thousands of craters on small protocontinents, superficially resembling the cratered surfaces of Mercury and the Moon. The crater basins on Earth, unlike those on the Moon and Mercury, filled with water and cosmic building blocks, developed a complex system of subsurface hydrothermal vents likely to be crucial to the origin of life. They provided both chemicals and energy sources that made hydrothermal crater vents a potential habitat for the origin of life. Hydrothermal activity can be triggered by impacts on a water-rich planet like Earth or even Mars. Such high-energy, high-temperature events create underwater areas boiling with heat and spewing chemicals. Fresh water in the crater lakes would be the ideal solvent enabling the cosmic ingredients to initiate prebiotic synthesis.

Hydrothermal crater lakes are ideal habitats where hyperthermophiles (extreme heat-loving microbes) thrive today surrounded by superheated water. Hyperthermophiles—the most primitive living organisms—support the view that life began in scorching environments [19]. There are more than 150 impact craters on Earth in a wide diversity of microbial habitats, ranging in size from the ~1.8-km-diameter Lonar Lake structure in India to the ~250-km-diameter Sudbury structure in Canada [7, 8]. Depending on their size, impact hydrothermal systems last for a finite period of time as the crater cools. The larger the crater, the longer the residence time. On young Earth, closely spaced craters of different sizes were likely to be interconnected by networks of subsurface fractures and exchanging heat and chemicals, possibly for life processing [5]. Simple prebiotic compounds could form in the extremely hot stage of large craters, whereas smaller craters would be conducive to more complex molecules. Small hydrothermal impact crater lakes appear to be the most plausible environments for the prebiotic synthesis of life on early Earth and even on Mars. Therefore, the National Aeronautics and Space Administration's (NASA) 'Curiosity Rover' has been exploring the 3.8-billion-year-old Gale crater of Mars for evidence of early life [23].

A single hydrothermal crater lake provides different habitats, such as a highly fractured central uplift with spewing chemicals, impact ejecta deposits, an annular basin, a crater rim for sequestering chemical reactions, postimpact water sediments on the lake floor, and diverse temperature gradients in the water column, where many chemical processes can co-occur (Fig. 4.1). Molten rocks from the central peak may heat cold freshwater of the crater lake to hot water, driving the convection of lake water

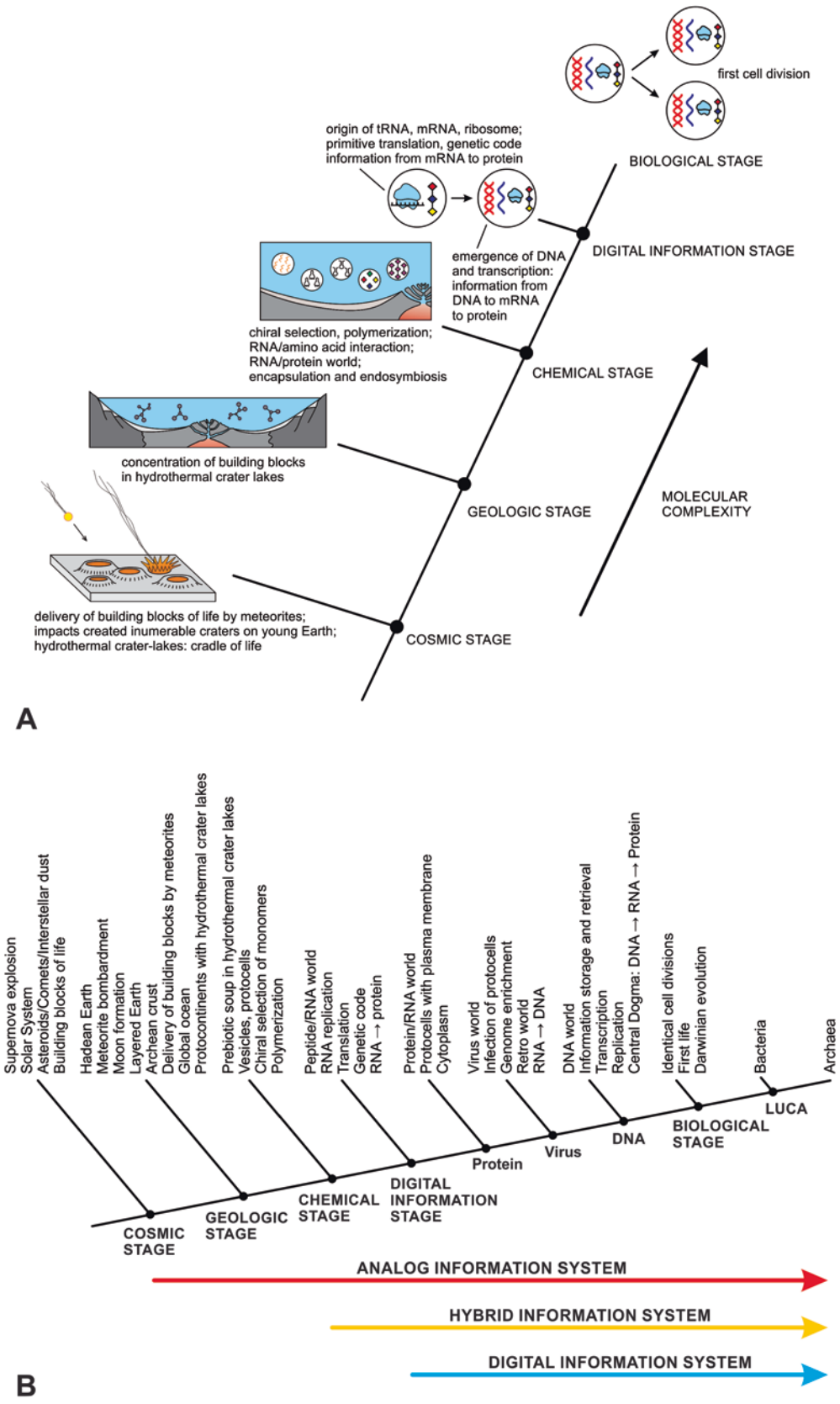


Fig. 4.1 (**a**) The hierarchical origin of life, viewed as five ascending stages of increasing complexity, showing the biomolecules in the prebiotic world that led to the development of the first cells. These are the cosmic, geological, chemical, digital information, and biological stages—each higher level acquired novel emergent properties. In the dark, hot environments of hydrothermal crater lake basins, prebiotic synthesis led to the creation of the first life. (**b**) The three ways of processing information in life are analog, hybrid, and digital, shown against the hierarchy of life

like a giant cooking pot. According to our hypothesis, cosmic building blocks, delivered by impacts, began to accumulate in the crater lakes, where hydrothermal energy would drive the synthesis of ever more complex organic compounds. The prolonged convective circulation of heated water, the temperature gradient, and plenty of carbon helped the mixing and concentration of organic molecules and increased their chemical activities. The catalytic mineral surface on the floor of the crater basin may have joined monomers and condensed them into larger complex polymers [24]. It should be noted, however, that not all the products of cosmic chemistry were used to construct life. Many molecules were discarded. Out of the vast numbers of diverse cosmic molecules, only a few, those with molecular compatibility and affinities to form molecules of greater complexity, would be selected for life-building processes.

Several critical components derived from meteorite impacts—lipid membranes, amino acids, sugars, phosphates, and nucleotides—began to interact in the crater lakes in the geological stage. These cosmic molecules, big and small, carried out a specific function in the vent environment using specific information. Each molecule had its own information system embedded. It performed a specific function based upon its structure and information. 'Analog information' emerged to guide the interaction of the building blocks of life in an orderly and organized manner in the vent environment. The critical molecules generated in hydrothermal crater vent environments were primitive lipid membranes. As we have noted, a membrane is the defining boundary of a cell, the basic unit of life. Lipid cell membranes were first formed from cosmic ingredients by self-assembly, floating on the lake's water surface like a thick oil slick [11]. Lipid membrane vesicles were also formed and crowded on the mineral surfaces like blisters. These membrane-bounded components began to encapsulate various biomolecules and monomers, separating them from their surroundings and enhancing interaction probability. Compartmentalization was necessary for the earlier stages of the prebiotic synthesis of protocells. Such encapsulation offered many potential benefits to the emergence of protocells of increasing complexity. In the early days of abiogenesis, primitive cell membranes began to encapsulate various organic molecules from the vent environment to form primordial protocells floating on the crater lake's surface. Encompassing these components, these protocells were tiny compartments now protected by membranes from the outside environment, each a minuscule natural laboratory for the chemical syntheses to follow. From geology, the chemistry was born.

4.5 The Chemical Stage

In the 'chemical' stage, hierarchical complexity starts to get truly serious. Geochemical processes, in conjunction with cosmic delivery of biomolecules present on early Earth, must have defined not only the types of molecules present in the hydrothermal crater lake environment but also the molecular composition of early chemical systems and, by extension, that of the protocells. Furthermore, the environmental conditions of hydrothermal vents must have defined the potential reactivity of the compounds, such as hydrogen cyanide (HCN) and formaldehyde (HCHO), which readily undergo further reaction in aqueous solutions. The combinatorial chemical reactions of the cosmic building blocks started to produce more complicated chemical species. These then catalyzed sets of more complicated results [11]. The water in hydrothermal crater lakes would have been abundant with potential electron acceptors (anions of various oxides), promoting avid recombination. High-temperature interactions between fluids and solids formed reduced gases (H_2S, hydrogen (H_2), and methane (CH_4)), which dissolved in hydrothermal fluids. Various mixtures of organic molecules were continuously recombined to create complex interacting systems and then exposed to sources of energy available in the hydrothermal vent environments, such as heat, oxidation–reduction potentials, and adenosine triphosphate (ATP) [17].

The chemical stage encompasses the plausible physicochemical mechanisms by which the protocells enclosing genetic and catalytic precursors could have assembled themselves. Today, all life is homochiral and uses only left-handed amino acids (L-amino acids) and right-handed sugars (D-sugars). In a vent environment, both amino acids and sugars produced chiral molecules in 50:50 mixtures. In the early stage of chemical evolution, L-amino acids and D-sugars were selected from heterogeneous populations by crystal faces of certain common minerals such as calcite, feldspar, quartz, and diopside [24]. The homochirality of biological molecules, L-amino acids, and D-sugars is a signature of life. Chiral properties provide an extra dimension for molecular recognition: between the enzymes and their substrates or between the amino acids and the RNAs. Simple chiral organic molecules such as amino acids and nucleotides could associate with forming polymers. One amino acid can join with another, creating a peptide bond, and two nucleotides can participate by a phosphodiester bond. The repetition of these reactions led to linear polymers known as polypeptides and polynucleotides. Nucleotides and amino acids, monomers for RNAs and polypeptides, were selected from random assemblies of molecular pools and then polymerized on the pores and pockets of the mineral substrate to create RNAs and polypeptides [24].

Polymers such as polypeptides and RNAs are central to living processes. Condensation or loss of water (dehydration) between two monomer molecules facilitates linking them to form long chains of polymers. The likely dehydration mechanisms during the polymerization of these monomers include both mineral surfaces on the crater floor [24] and a repeated wet-and-dry cycle on the water surface of the crater basin [11]. The catalytic nature of the crater floor's mineral surface may have also played an essential role in the polymerization of RNAs and peptide molecules. Several mineral catalysts for polymerization, such as montmorillonite clay and pyrites, abounded in the hydrothermal vent environments. Empty lipid membranes began to encapsulate various monomers and polymers for efficiency, stability, and molecular symbiosis.

Mineral catalysts may have played essential roles in establishing early metabolism. Apatite might have helped in the building of the cell membrane because of its phosphate content. Metal ions of iron, manganese, zinc, and copper were also available in the vent environment, which would help mediate catalysis [24]. Crystalline surfaces of common rock-forming minerals such as pyrite and montmorillonite on the crater floor enhanced protometabolism by polymerization of nucleotides into RNAs and amino acids into peptides. Complex molecules, such as RNAs, polypeptides, and phospholipid membranes, would have originated from small molecules whose reactivity was guided by physicochemical conditions. ATP's chemical energy was used to remove the equivalent water molecules from the amine and carboxyl groups on amino acids to form peptides bonds or from the phosphate and ribose hydroxyl groups on nucleotides to form ester bonds [17]. Self-assembly of membranous compartments encapsulated these polymers, peptides, and RNAs into populations of protocells. In prebiotic chemistry, both condensation and self-assembly played crucial roles in creating more and more complex molecules. Natural selection at the molecular level, creating increasing complexity under prebiotically possible conditions, emerges as a decisive factor in successive stages of biogenesis (Fig. 1.2). This may explain why living systems exceed nonliving structures in sheer complexity, and so few abiotic precursors of living systems would survive the early stages of prebiotic synthesis.

Regarding prebiotic precursors, since the discovery of the catalytic RNA molecule, called the 'ribozyme,' the 'RNA world' model of the prebiotic environment has become the leading paradigm in current origin-of-life research [25–28]. However, in recent times, what has been called the peptide/RNA world appears to be a more parsimonious theory than the RNA world per se. We believe that there was a molecular symbiosis between peptides and RNAs in protocells. In the prebiotic environment, the synthesis of peptides, formed by linking amino acids, was relatively easier to accomplish than the polymerization of nucleotides to RNAs. There is thus no justification for excluding peptides from prebiotic chemical reactions. Ribozymes and peptide enzymes working together could catalyze chemical reactions more efficiently than ribozymes working alone. Several authors support the peptide/RNA world model on the grounds that a single polymer like RNA could not carry out all of the necessary processes for the emergence of the genetic code and protein synthesis [29–31]. Various peptides in the prebiotic vents were available to assemble the first genetic codes and the translation machinery to carry out necessary protein synthesis processes. Since the RNA world alone could not achieve this outcome without the active collaboration of peptides, the chemical stage most likely produced a peptide/RNA world.

In the peptide/RNA world, various noncoding RNA molecules such as ribozymes, pre-tRNA, bridge peptides, and ribosomes appeared that would make important components of the translation machine. This is the age of the hybrid information system (HIS). Some essential functions, including those in the translation, are performed by noncoding RNA molecules. Pre-tRNA and tRNA would create mRNA step by step to usher in the digital information system (DIS). From chemistry, the digital information stage was born.

4.6 The Digital Information Stage

The information found in genes is digital and widely known. Digital information is encoded in linear nucleotide polymers such as mRNA and DNA. We know that any computer information is partly digital and partly analog, and any living organism is an even more tangled mixture of both digital and analog components. The concepts of analog and digital are far too narrow to encompass the subtilities of living cells because, in many cases, the information in a living system is combinatorial. Here, the analog vs. digital dichotomy is more of an analogy than an accurate description. Yet, it provides working tools to investigate the origin and evolution of information systems in life and how these two processes interact, often giving rise to one another.

Analog information is present in various chemical components that pass from generation to generation and vary from cell to cell. It began during the onset of abiogenesis. In the cosmic stage, primitive Earth favored the analog format to begin prebiotic synthesis. In analog chemical systems, information is contained in a continuous variable composition of an assembly of molecules in a prebiotic soup rather than in a discrete string of four digital bases (Fig. 4.1b).

The dynamic molecular environment of the chemical stage in the peptide/RNA world gradually yielded a coherent, interconnected network of chemical processes—the 'digital information' stage. Lipid membranes encapsulated amino acids, RNAs, and peptide molecules drawn from the prebiotic soup in the vent environment to initiate molecular

symbiosis inside the protocells. This prebiotic endosymbiosis led to the hierarchical emergence of several critical components of the translation machine: transfer RNAs (tRNAs), aminoacyl-tRNA synthetases (aaRSs), ribozymes, various peptide enzymes, and finally ribosomes. At this stage, the carrier of the information system, the messenger RNA (mRNA), was built step by step to encode information, the recipe of proteins. The synthesis of proteins occurred simultaneously with the emergence of a translation system and the genetic code. This was the beginning of the digital information age. Translation and the genetic code developed side by side. The translation machine decoded the message from mRNA to create polypeptide chains or proteins.

The climax of the digital information stage was the advent of analog protein chains for coming up with several crucial biochemical innovations, including the emergence of the plasma membrane, the primordial cytoplasm, gene regulation, the virus world, and DNA. The newly synthesized proteins heralded the virus world when proteins coated mRNA molecules in the vent environment for durability and protection. The first branching in abiogenesis was when the virus world and the protocellular world evolved side by side, interacting with each other. Primitive RNA viruses began to infect protocells and exploit their ribosomes for reproduction. The next stage in viral evolution was the emergence of the first retrovirus with a new kind of replicative strategy. It could facilitate the transition from RNA to DNA genomes using its reverse transcriptase enzyme. With persistent infection, DNA viruses slowly transferred not only their core replication enzymes but also their DNAs. Thus, emerged the DNA world when DNA replaced RNA as the major genome of the protocells [32, 33]. Genetic information began to flow from DNA to mRNA to proteins in a two-step process involving both transcription and translation. The advent of DNA entirely dissociated the replication of information from its expression by translation. Because DNA is much more stable than mRNA and has more storage capacity, it is a superb archive for information systems inscribed in the form of base sequences. DNA progressively took over the replicative storage function of mRNA, reserving the latter for protein synthesis. From the digital information stage, biology was born with the emergence of first life.

4.7 The Biological Stage

With the arrival of the 'biological' stage, all cell components could now cooperate symbiotically in a unitary production with a reproductive capability. Life emerged from the capacity of the first cells to process information coded in DNA to direct the cells' self-production and reproduction. DNA replication was central to the first cell's binary fission, orchestrated by the duplication of genomes and then dividing the parent cell into two identical daughter cells. With the onset of binary fission, the population of primitive cells increased in hydrothermal vent environments, undergoing Darwinian evolution and diversification by mutation and recombination. The status of 'life' was established when the first cells maintained and repaired themselves and reproduced and evolved. The reproduction phase involves replicating all internal content (analog and digital information content) within a chemical system that could enlarge its compartmental boundaries. The maintenance growth stage prepared for a division–reproduction event in which a parent cell divided into two identical offspring.

4.8 Conclusions

Hierarchy is a universal organizing principle in biology and a key reason molecular evolution produced complex, evolvable first cells. We have applied recent astrobiological research to our scenario in which life arose about four billion years ago through five hierarchical stages of increasing molecular complexity in a steaming hot environment of hydrothermal crater basins.

In the 'cosmic' stage, a star explosion near the solar nebula cast the building blocks of life into interstellar space. During the Late Heavy Bombardment period, the comets and carbonaceous chondrites produced within that nebula transported water and organic molecules to young Earth. Asteroid and meteorite collisions created numerous hydrothermal crater lakes on the Archean crust, crafting cradles for prebiotic chemistry.

In the 'geological' stage, crater basins containing an assortment of cosmic and terrestrial organic compounds powered by hydrothermal and chemical energies drove the early processes of prebiotic synthesis. The analog information system (AIS) emerged to move and concentrate the building blocks in the vent environment. It created lipid membranes that floated as a thick oil slick upon the crater basin's watery surface.

In the 'chemical' stage, nucleotides and amino acids were selected from random assemblies of the prebiotic soup and polymerized at the pores of mineral surfaces, giving rise to peptide and RNA polymers. These biopolymers were randomly encapsulated to initiate a molecular symbiosis, thus furthering the hierarchical emergence of complex protocell components. In the peptide/RNA world, several noncoding RNA molecules such as pre-tRNA, activating enzymes such as bridge peptides, and ribosomes originated from ribozymes to from components of the translation machine. This is the age of the hybrid information system (HIS) that would give rise to the digital information system.

In the 'digital information' stage, pre-tRNA and tRNA became molecular architects for designing and encoding

pre-mRNA and mRNA step by step, creating the genetic code as a memory bank. The newly synthesized proteins from the nucleotide sequences of mRNA led to the emergence of RNA viruses. As RNA viruses infected RNA protocells, retroviruses evolved. Retroviruses, in turn, created DNA and donated DNA to protocells during infection. DNA in protocells became the primary genetic storage system. Digital information began to flow from DNA to RNA to proteins.

The critical breakthrough to life arrived in the biological stage with the emergence of the first cells capable of self-production, reproduction, heredity, variation, and Darwinian evolution. The origin of the first cells was an event horizon that transformed the rocky, sterile world into a living, evolving world.

References

1. Dodd MS, Papineau D, Grenne T, Slack JF, Rittner M, Prajno F, O'Neil JO, Little CTS (2017) Evidence for early life in Earth's oldest hydrothermal vent precipitates. Nature 543:60–65
2. Mojzsis SJ, Harrison TM (2000) Vestiges of a beginning. Clues to the emergent biosphere are recorded in the oldest known sedimentary rocks. GSA Today 10:1–5
3. Cockell CS (2015) Astrobiology: understanding life in the universe. Wiley Blackwell, Sussex
4. Bernstein MP, Sandford SA, Allamonda LJ (1999) Life's far flung raw material. Sci Am 263:42–49
5. Chatterjee S (2016) A symbiotic view of the origin of life at hydrothermal impact crater lakes. Phys Chem Chem Phys 18:20033–20046
6. Chatterjee S (2018) The hydrothermal impact crater lakes: the crucibles of life's origin. In: Kolb VM (ed) Handbook of astrobiology. CRC Press/Taylor & Francis, Boca Raton, pp 265–295
7. Cockell CS (2006) The origin and emergence of life under impact bombardment. Philos Trans R Soc B361:1845–1856
8. Osinski GR, Tornabene LL, Banerjee NR, Cockell CS, Flemming R, Izawa MRM, McCutcheon J, Parnell J, Preston LJ, Pickersgill AE, Pontefract A, Sapers HM, Southam G (2013) Impact-generated hydrothermal systems on Earth and Mars. Icarus 224:347–363
9. Boss AP, Keiser SA (2010) Who pulled the trigger: a supernova or an asymmetric giant branch star? Astrophys J Lett 717:L1–L5
10. Delsemme AH (1988) Our cosmic origins. Cambridge University Press, Cambridge
11. Deamer DW (2011) First life: discovering the connections between stars, cells, and how life began. University of California Press, Berkeley
12. Pizzarello S, Cronin JR (2000) Non-racemic amino acids in the Murchison and Murray meteorites. Geochem Cosmochem Acta 64:329–338
13. Chyba C, Sagan C (1992) Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origin of life. Nature 355:125–132
14. Baross JA, Hoffman SE (1985) Submarine hydrothermal vents and associated gradient environment as sites for the origin and evolution of life. Orig Life 15:327–345
15. Martin W, Russell MJ (2007) On the origin of biochemistry at an alkaline hydrothermal vent. Philos Trans R Soc B362:1887–1926
16. Martin W, Baross J, Kelley D, Russell MJ (2008) Hydrothermal vents and the origin of life. Nat Rev Microbiol 6:805–814
17. Martin WF, Sousa FL, Lane N (2014) Energy at life's origin. Science 344:1092–1093
18. Farmer JD (2000) Hydrothermal systems: doorways to early biosphere evolution. GSA Today 10:1–9
19. Parnell J, Lee P, Cockell CS, Osinski GR (2004) Microbial colonization in impact-generated hydrothermal sulphate deposits, Haughton impact structure, and implications for sulfate on Mars. Int J Astrobiol 3:247–225
20. Mulkidjanian AY, Bychkov AB, Diprova DV, Galperin MY, Koonin EB (2012) Origin of first cells at terrestrial, anoxic geothermal fields. Proc Natl Acad Sci U S A 109:E821–E830
21. Bernhardt HS, Tate WP (2012) Primordial soup or vinaigrette: did the RNA world evolve at acidic pH? Biol Direct. www.biology--direct.com/content/7/1/4
22. Tang M, Chen K, Rudnick RL (2016) Archean upper crust transition from mafic to felsic marks the onset of plate tectonics. Science 351:373–375
23. Grotzinger JP et al (2014) A habitable fluvio-lacustrine environment at Yellowknife Bay, gale crater. Mar Sci 343. https://doi.org/10.1126/science.1242777
24. Hazen RM (2005) Genesis: the scientific quest for life. Joseph Henry Press, Washington, DC
25. Gilbert W (1986) The RNA world. Nature 319:618
26. Cech TR (1986) RNA as an enzyme. Sci Am 255(5):64–75
27. Orgel LE (1994) The origin of life. Sci Am 271(4):77–83
28. Joyce GF (2002) The antiquity of RNA-based evolution. Nature 418:214–221
29. Carter CW (2015) What RNA world? Why a peptide/RNA partnership merits renewed experimental attention. Life 5:294–320
30. Carter CW, Wills PR (2017) Interdependence, reflexivity, impedance matching, and the evolution of genetic coding. Mol Biol Evol. https://doi.org/10.1093/molbev/msx265
31. Chatterjee S, Yadav S (2019) The origin of prebiotic information system in the peptide/RNA world: a simulation model of the evolution of translation and the genetic code. Life 9:25. https://doi.org/10.3390/life9010025
32. Forterre P (2006) The origin of viruses and their possible roles in major evolutionary transitions. Virus Res 117:5–16
33. Koonin EV, Dojla VV (2013) A virocentric perspective on the evolution of life. Curr Opion Virol 3(5):546–557

Cosmic Connections

5

All of the rocky and metallic material we stand on, the nitrogen in our DNA, the calcium in our teeth, the iron in our blood, the carbon in our genes were produced billions of years ago in the interior of a collapsing star. We are made of star-stuff.

—Carl Sagan, 1990

5.1 Supernova Explosion and the Origin of the Solar System

The origin of life encompasses a wide range of phenomena of cosmic origins. For decades, scientists have suspected that the explosive death of a small star – about 12 times heavier than our Sun – at the end of its life cycle—an event called a supernova—helped trigger our solar system's formation about 4.6 billion years ago. Stars are like giant hydrogen bombs in which nucleosynthesis is the powerhouse. This converts hydrogen into helium by nuclear fusion, and this transformation releases enormous atomic energy. The helium inside stars is converted into carbon, nitrogen, oxygen, iron, sulfur, and phosphorous—everything that life is made of. When the core of a giant star (somewhere between 8 and 15 times more massive than the Sun) reaches the stage of producing iron, it finally runs out of hydrogen fuel. The outer layers of the star will then collapse into the iron-rich core. This event triggers a supernova explosion. The cataclysmic annihilation of a dying star creates new, heavier elements and blasts most of its material out into the surrounding space [1, 2].

In our case, the shock wave from a core-collapsing supernova emitted stardust into the solar nebula, whose gravitational collapse set our solar system spinning (Fig. 5.1). The dying star shed much of its mass in silicate mineral dust that formed immense molecular clouds throughout our galaxy. The solar system thus developed from a cloud of dust and gas; it froze interstellar gas into ice and dust particles that now contain rich organic molecules. They would grow by collision and accretion. We believe that the solar system coalesced from the solar nebula, a giant rotating cloud of gas and dust that eventually contracted at its dense center to produce a new, smaller star, namely, the Sun, surrounded by its planetary system. A new star arose from the ashes of the dying star. The dust particles of its nebula aggregated into planetesimals that ultimately accreted into planets. After the Sun formed, the gas giants Jupiter and Saturn followed, and the rocky planets attained most of their present size. Today's asteroids and comets are planetesimals that escaped participation in planet formation. Eons ago, as the rocky planets, which were once planetesimals themselves, accreted circumambient materials, they also suffered increasingly violent collisions with asteroids, comets, and other planetesimals. One such immensely impactful event at the end of Earth's accretion phase is believed to have smashed off a piece of infant Earth while creating our relatively massive Moon.

The solar system consists of our star—the Sun—and everything gravitationally bound to it—all the planets and their moons as well as millions of asteroids, comets, and meteoroids. Energy from the Sun's deep hot core, generated by the thermonuclear fusion of hydrogen and helium, radiates outward in the form of sunshine. The journey of heat from the core to the surface takes more than 100,000 years. Small stars like the Sun will face a relatively peaceful and beautiful end after proceeding through several stages. However, massive stars are all headed toward a violent end in a glorious explosion—a supernova. A low-mass supernova 4.6 billion years ago triggered the formation of our solar system and its planets.

Meteorites preserve records of the elements, isotopes, and compounds derived from the stardust of the solar system's formation. Meteorite chemistry provides a fundamental constraint for testing the foregoing theory of life's cosmic seeding by the supernova explosion. Supernovae produce legible patterns of short-lived radionuclides, which would be delivered today in meteoritic samples as isotopic anomalies such as beryllium-10 [1]. To see whether a supernova could explain the pattern of isotopes observed in primitive meteorites, scientists at the Carnegie Institution in Washington, DC,

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_5

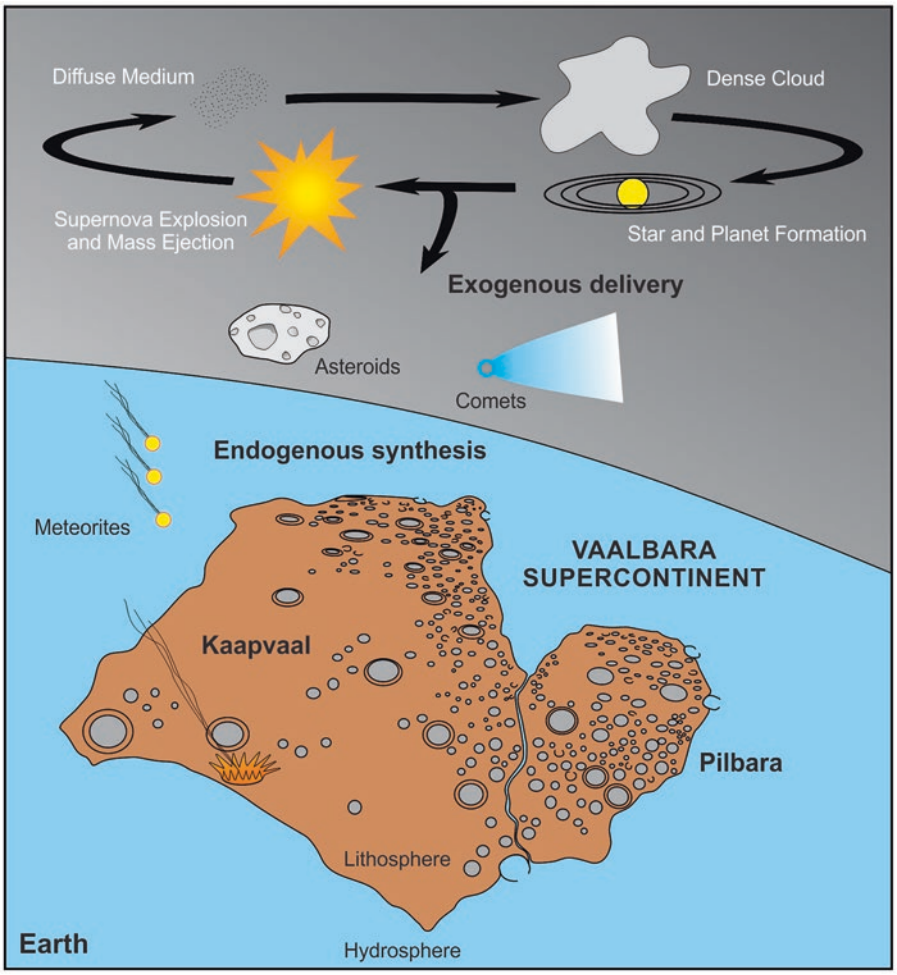


Fig. 5.1 From the outflows of a supernova explosion, reactions, and radiations of the interstellar medium, through nebular and solar system chemistry, cosmic building blocks were delivered to young Earth by meteorite impacts during the tail end of the Heavy Bombardment period (exogenous delivery). However, abiogenesis occurred in the hydrothermal crater lakes on the Archean protocontinent such as Vaalbara (endogenous production), leading to the creation of the first cells

developed computer models of supernova shock waves and solar system formations. These models indicated that the iron-60 in primitive meteorites probably came from supernova debris. Scientists have detected crucial evidence of this supernova explosion in the form of short-lived radioactive isotopes such as iron-60, trapped within the primitive meteorites. Since only a supernova can produce iron-60, we have tried to explain its origin by suggesting that a supernova occurred nearby, thus spreading the iron isotope around. The ratio of aluminum-26 to magnesium-26 isotopes in ancient meteorites is accorded a similar explanation [2].

One of the most striking discoveries made from the analysis of the Allende meteorite, which fell near Chihuahua, Mexico, in February 1969, was the crucial evidence that a nearby star exploded to form a supernova about 4.6 billion years ago [3]. The Allende meteorite is a carbonaceous chondrite that almost perfectly preserves materials left over from the solar system's very birth. Materials recovered from the Allende meteorite revealed compositions that must have derived from dying stars shortly before the development of the solar system and provided physical evidence of those processes. Therefore, we are confident in asserting that the building blocks of life arose as by-products of this star's supernova explosion, which were dispersed as interstellar dust, asteroids, and comets during the beginning of our solar system [4–7]. Newly manufactured atoms and molecules enriched interstellar space with the remains of the dying star. This interstellar material went into the creation of the Sun, its planets, and its meteorites. Furnaces deep inside ancient stars manufactured the atoms of iron and nickel that make up Earth as well as the carbon in our bodies and the oxygen we breathe. Carrying stardust in its DNA, Earth is the living ember of a dying star. Shiva is often, and most beautifully, presented in the form of Nataraja, the cosmic dance. He holds the flame of destruction in one hand, and a drum to regulate the rhythm of the dance (and symbolize creation) in another. He moves in a ring of fire—the cosmic cycle—maintained by the interaction of creation and destruction Like the cosmic dance of the Hindu god Shiva—the interplay between destruction and renewal—life on Earth was born from the death of a star.

The planets formed in the solar system differentiated into the inner, rocky terrestrial planets and the outer, gaseous Jovian planets. The inner planets—Mercury, Venus, Earth,

and Mars—are relatively small. With an abiding abundance of liquid water on its surface, the third planet from the Sun is the only place we know that living beings inhabit. With a radius of 6371 kilometers, Earth is the biggest of the terrestrial planets. It formed as a result of the condensation of cosmic dust and then grew from a small planetesimal until it reached its present size by the accretion of dust, meteorites, and other planetesimals. It may have taken more than a 100 million years for proto-Earth to grow from about 10 km in diameter to its current size. Bearing in mind the heat generated throughout the history of accretion, young Earth was a hellish world of scalding rocks and choking fumes as hot as volcanic lava. Composed mostly of water vapor, carbon dioxide, and nitrogen, its superheated atmosphere wore an opaque veil of clouds.

5.2 Asteroids and Comets

Spacecraft can now study asteroids and comets at a close range. Fortunately, pieces of asteroids that fall to Earth from time to time are available for scientific examination. Meteorites are asteroidal fragments that, along with cosmic dust, offer a glimpse into the early history of the solar system's formation. An estimated 40,000 metric tons of extraterrestrial material fall to Earth every year, mostly as interstellar dust rather than meteorites. A few meteorites probably represent pieces of the Moon or Mars that were blasted off their surfaces by asteroidal impacts and eventually landed on Earth. Scientists and collectors alike prize meteorites for their rarity and exotic provenance. Sometimes known as 'the poor man's space probe,' these remarkable extraterrestrial objects require interrogation by sophisticated and expensive instruments to reveal their ancient secrets about the origin of the solar system, planetary water, and early life.

A belt containing millions of asteroids lies between the orbits of Mars and Jupiter. Fragments of asteroids rain down on the inner solar system as meteorites. Asteroids are rocky airless remnants left over from the early formation of our solar system about 4.6 billion years ago. They range in size from Ceres—about 940 km in diameter—to bodies that are less than 2 m. Most asteroids fall into three categories based on their composition:

1. The grayish C-type or carbonaceous asteroids are the most common. They include more than 75% of known asteroids and mainly consist of clay and stony silicate rocks. Inhabiting the central belt's outer region, they are rich in organic compounds and volatiles and are closely linked to the origin of our planet's life and oceans.
2. The greenish to reddish S-type or silicaceous asteroids represent 17% of the known asteroids. Consisting of silicate material and nickel–iron, they dominate the inner asteroid belt.
3. The reddish M-type or metallic asteroids make up the rest of the asteroids. They occupy the central region of the main asteroid belt and are composed of nickel–iron.

Meteorites are themselves of several types, depending on the relative amounts of silicate minerals and iron that they contain: stony (from C-type asteroids), stony-iron (from S-type asteroids), and iron (from M-type asteroids). Stony meteorites are the most abundant and are of two types: chondrites and achondrites. Carbonaceous chondrites contain carbon, whereas achondrites do not. The Murchison meteorite, which was dated about 4.5 Ga to determine the age of Earth, is a carbonaceous chondrite. Carbonaceous chondrites may have been an essential source of the building blocks of life and water on Earth.

Chondrites are sterile stony meteorites that have not been modified by the melting or differentiation of the parent body. They are the most prevalent type of meteorite to fall to Earth, representing more than 75% of all asteroids [8]. These asteroids provide critical information on the origin and age of the solar system, the synthesis of the building blocks of life in interstellar dust, the origin of life, and the presence of water on Earth. Like most asteroids, chondrites originated from the asteroid belt, where collision and gravitational perturbation put them into an Earth-crossing orbit. Among chondrites, carbonaceous chondrites are so primitive that they contain traces of interstellar dust, including organic compounds that survived the thermal processing of the solar nebulae. The carbon compounds they carry are the primary components of life. Of all chondrites, they formed farthest from the Sun with the highest proportion of volatile compounds and water. Carbonaceous chondrites may be relatively unmodified material from the solar nebulae. These meteorites are 85%–90% hydrated silicates of iron and magnesium and 3%–4% carbon and carbon compounds. Microscopic examinations of C chondrites show various dust grains of different compositions, gently compressed together like a sedimentary rock. These carbonaceous chondrites originated from the accretion disk that surrounded the nascent Sun. Their collision with Eoarchean Earth probably delivered crucial organic components and water. Hence, their study provides important clues for understanding organic synthesis, the origin of life, and the presence of water on Earth.

Unlike asteroids, comets are 'dirty snowballs,' relics from the dawn of the solar system around 4.6 billion years ago, made of ice and frozen ammonia, methane, and carbon dioxide, with some dust and organic molecules thrown in. When a comet is proximate to the Sun during its orbital path, it thaws out and spews dust and gases to form a giant glowing head like a fuzzy blob with a long tail pointing away from the Sun. There may well be billions of comets orbiting our Sun

in a doughnut-shaped ring, known as the Kuiper Belt, extending beyond Neptune and the even more distant Oort Cloud.

The chemical composition of asteroids and comets provides critical information about the raw materials of the organic compounds present in interstellar space, which are of prebiotic interest. While meteorites were more likely to collide with early Earth, being much more numerous in the Hadean and Late Heavy Bombardment periods (~4.1–3.8 Ga) than they are today, asteroids were more likely than were comets to have triggered the impact cataclysm during this Late Bombardment period, thus affecting the entire inner solar system [9]. The asteroid projectiles dramatically altered surface conditions in the Earth–Moon system, as well as on Mercury and Mars, and seeded the inner planets with organic compounds and water from the very beginnings of our solar system. They were the 'manna from heaven' to sterile Earth.

5.3 The Hadean Eon: The Origin of Earth and the Moon

In her infancy, Earth was not as beautiful as she is today. She had a traumatic early childhood in a hostile environment. Asteroids and comets regularly crashed into the planet ever since it formed, especially during its earliest history. This was the Hadean Eon (~4.6–4 Ga). Its geological record is fragmentary, but the impact history preserved on the lifeless Moon and the inner planets' pockmarked surfaces help us in reconstructing that period. Throughout the Hadean Eon, the impact of extraterrestrial bodies released enormous heat that probably prevented much of the rock from solidifying on the surface. Impacts and collisions were the dominant geological processes during the Hadean and early Archean periods, playing a vital role in the crustal, tectonic, thermal, magmatic, and environmental evolutions of young Earth. During its 500 million years, Hadean Earth was violent and desolate, enduring the punishing rain of extraterrestrial impacts that scorched the globe and widely reprocessed the crustal surface through the mixing and burial of impact-generated melt.

At first, Earth was a barren planet, inhospitable to living organisms, and a seething cauldron of erupting volcanoes, raining meteors, and hot noxious gases in a steamy atmosphere. Comets and asteroids continuously battered young Earth and the other inner planets for the first 600 million years. Hadean Earth was a fiery globe of molten crust generated by the relatively heavy bombardment of meteorites. It was a black, bleak, and burning planet with no signs of life. One can picture Dante's inferno: the atmosphere was thick and opaque to thermal radiation while internal heat dominated the climate [10]. Hadean Earth differed from living Earth in at least two significant ways: (1) extraterrestrial bolides, both large and small, continuously bombarded the planet and (2) with its ocean of magma, it had significantly more thermal energy than it has today. Our planet's impact record and violent beginnings have since been erased from its surface by volcanic resurfacing, active erosion, plate tectonic activity, and the moderating influences of the biota. Most but not all the organic compounds present during the planet's accretion from interstellar dust were broken down during the molten stage of Hadean Earth.

5.4 Moon-Forming Impact and Iron Catastrophe

Both waxing and waning, the phases of the Moon, have been an object of fascination and wonder in every culture. Our Moon is unusual among solar system satellites because of its large mass. It comes in at more than 1% of the mass of the planet it orbits, i.e., 1.2% to be precise. Did the Moon form from a cataclysmic collision of two planets? This is the current consensus. Shortly after Earth's formation, a Mars-sized planet called 'Theia' smashed into it [11, 12]. A magma ocean might have covered Earth before the Moon-forming impact. This massive collision ejected an enormous amount of molten lava, gas, and debris from young Earth to create a disk that coalesced into the Moon by gravitational accretion (Fig. 5.2). As the material fused, it would have entered Earth's elliptical orbit around the Sun, held in place by the mutual gravity of the new Earth–Moon system. On the completely molten early Moon, the flotation of mineral feldspar (plagioclase) on the top of this magma ocean gave rise to lunar highlands, made of light-colored anorthosites. The crust of these lunar highlands is the original crust formed from that magma ocean, which extended to depths of about 500 km. Unlike Earth, whose core is 32% of its mass, the Moon has little or no core. Yet, the Moon's chemical and isotopic similarity to silicate Earth argues that it was predominantly formed from Earth's surface materials rather than the impactor Theia.

The impact hypothesis also accounts for the inclined orbit of the Moon and the angular momentum of the Earth–Moon system. After its birth, the Moon would have been close enough to have filled much of the sky of rapidly rotating Earth and to have exerted large tidal forces that slowed down Earth's spin, a process that continues even today. As the Moon moved farther away, the angular momentum of the total system was conserved. Theia left few chemical fingerprints on the Moon; rather, Theia's metallic core augmented that of Earth, bringing it up to its present mass.

Laboratory tests of Moon rocks from the Apollo series of explorations support this collision scenario [11]. Chemical data from the Apollo samples suggested a remarkable dwindling in water, sodium, and other volatile components compared with Earth rocks. The impact origin of the Moon accounts for its composition of predominantly lighter elements. It is less dense

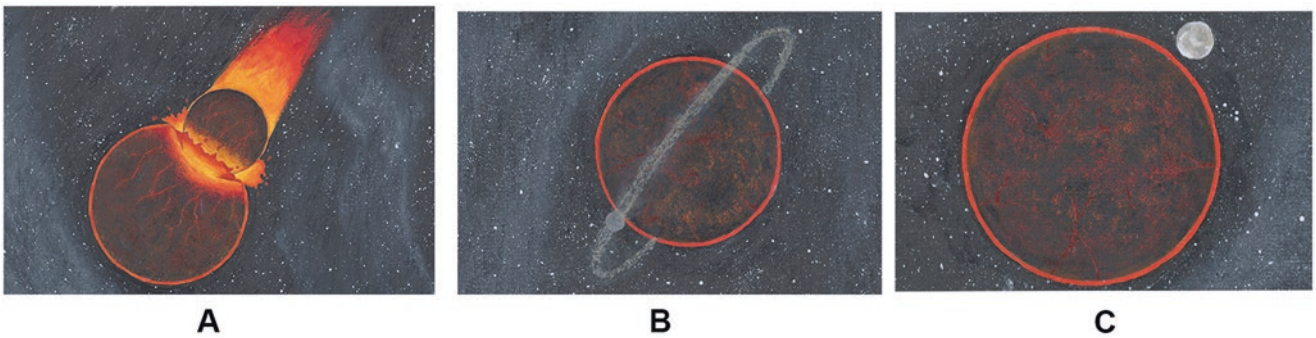


Fig. 5.2 (**a**) The inner solar system's most significantly known collision was the Moon-forming giant impact between a Mars-sized planet named Theia and Earth, approximately 4.5 billion years ago, about 30–50 million years after the origin of the solar system. (**b**) This impact created a debris ring that coalesced to form the Moon. (**c**) The Earth–Moon system, when the Moon was closer to Earth during the Hadean period and would fill a large area of the sky at the time

than Earth because the material that formed it came from Earth's crust and not its core. The surface of the Moon exhibits the following features that yield clues to its impact origin:

1. Rocks from Earth and the Moon have identical oxygen and zinc isotopic ratios.
2. The lunar crust's highly anorthositic composition suggests that a large portion of the Moon was once molten.
3. Earth's spin and the Moon's orbit have similar orientations.
4. Earth has the highest density of all planets in the solar system.

Because of this extremely early catastrophic event, Earth must have lost into space practically all the water and most of the organic elements (carbon, hydrogen, oxygen, nitrogen, phosphorus, and sulfur) that the protoplanet had previously retained. Comets and asteroids later captured by primitive Earth after the giant impact period must have been the primary source of the terrestrial volatiles and organic compounds needed for biosynthesis.

The heat generated by the Moon-forming impact triggered large-scale melting of Earth, forming an early magma ocean and an opaque atmosphere. Swirling currents churned Earth's interior. Meanwhile, radioactive minerals trapped deep in the interior generated immense heat. The 'iron catastrophe'—perhaps the most important episode in our planet's history short of the origin of life—occurred when early Earth's temperature exceeded the melting point of iron (1538 °C). Gravity slowly pulled the heavy liquid iron—the planet's most abundant element, i.e., about 35% of Earth's overall composition—toward the interior to form a core. Lighter molten compounds, rich in silicon, oxygen, and other light elements such as aluminum, sodium, calcium, and potassium, rose toward the surface and differentiated into three layers: the core, the mantle, and the crust. The 'iron catastrophe' formed a dense metallic core mostly of iron, a medium-density mantle, and a low-density crust. At 5700 °C, this iron core is as hot as the Sun's surface. Here, the flow of liquid iron generates electric currents, creating a magnetosphere that protects Earth's surface from deadly solar flares and whisks away their charged particles. The magnetic field causes compass needles to point toward the north pole, regardless of which way we turn. Gases that were not at least partly soluble in the melted material, such as water vapor and CO_2, formed a thick, reddish-tinged atmosphere around postimpact Earth [13].

The immense collision that formed the Moon spun Earth and tilted its axis of rotation by about 23°, thus introducing the day–night cycle and changing the wind patterns and seasons. Days on primeval Earth had perhaps 3 h of sunlight and 3 h of night. Bound by Earth's gravity, the Moon's orbit slowed Earth's rotation, thus lengthening the day and setting the stage for the arrival of oceanic tides. The Moon's own gravity affects Earth's liquid envelope and ocean tides. It also helped stabilize our planet so that its axis of rotation persists in the same direction and its climate remains stable over long time scales. Being so close to Earth, the moon generated massive tides compared to today. Those tides gave rise to wet/dry cycles at many surface spots in a regular manner. The wet/dry cycles are important for polymerization of nucleotides and amino acids, as discussed in Chap. 8 as well. That is one peculiar link between the origin of life and the Moon.

The Moon's record of crucial information strongly informs our current understanding of Earth's own development. Scars on the Moon's surface record the impacts it absorbed during the Late Heavy Bombardment. Impact craters formed at that time still dominate the surfaces of the Moon and Mercury. Both bodies lack liquid water that would have eroded those craters over time. During the

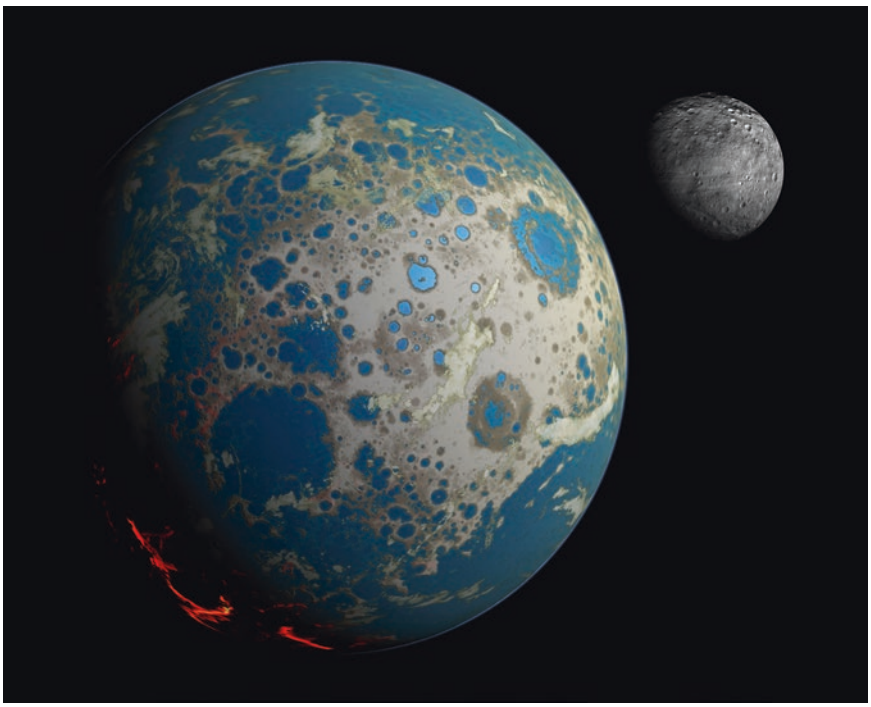


Fig. 5.3 An artistic concept of the early Earth–Moon system (~4.2 Ga). Reconstruction of the highly cratered surface of the Hadean Earth. Impacts continually reprocessed the surface of Hadean/Archean Earth, which mixed and buried the impact-generated melt. The Moon is shown as a dry, heavily cratered body, far less geologically active than Earth because it lacks plate tectonics and a dynamic atmosphere. The crater records of the Moon calibrate the Late Heavy Bombardment period on Earth. (Courtesy of Simone Marchi)

Late Heavy Bombardment, our planet would have resembled the Moon and Mercury. Unlike these planetary neighbors, however, water filled up the craters on Earth to form hydrothermal crater lakes. By now, plate tectonics, the continuous dynamics of the hydrosphere, wind, and other erosional forces, have erased all these impact craters from the face of Earth.

In contrast to the active geology of Earth, driven from below by its massive core, the Moon's surface is geologically dead. Crumbled rock called regolith, formed from the disintegration of basaltic and anorthositic rock under continuous meteoritic bombardment, covers its surface. It is a one-plate world, bereft of plate tectonics. Moreover, the Moon has no atmosphere: no weathering or erosion has scoured its surface. The Moon has hardly altered, maintaining its pristine state over the eons. We have this stasis to thank for preserving the early history of the inner solar system and the terrestrial planets, thus compensating to some degree for the missing record on our own planet. Clues obtained from the Moon, as well as from Mercury, Mars, meteorites, and computer simulations, have helped us to reconstruct the Hadean record. The erasure of Hadean rocks from Earth's geological record is due to their destruction by meteoritic bombardments that continued throughout the Hadean and early Eoarchean times (Fig. 5.3).

5.5 After the Moon's Formation

After the giant impact with Theia, Earth's surface cooled rapidly—that is, within a few million years—from rock-melting temperature (~1700 °C) to below the boiling point of liquid water. The cooling of the magma ocean created the first planetary crust. After the Moon-forming event, interplanetary dust particles (IDPs) became the main source of mineral mass continuously delivered to Earth. The Hadean crust was probably mafic or basaltic for the most part like that forming the ocean floors. Felsic or broadly granitic rocks are the defining features of the continental crust. No crustal rocks have survived from the time of intense Hadean bombardment, between Earth's formation 4.6 Ga ago and the oldest known components of the Acasta Gneiss in northwestern Canada, dated at just over 4 Ga. The only valid evidence of the Hadean crust has come from the only known survivors of the Late Heavy Bombardment—zircon crystals from Australia and India, where surface water played an essential role in their formation [14].

The Detrital zircons provide evidence of a Hadean crust in a chert–pebble conglomerate in the Jack Hills in the Yilgran Craton of Western Australia, where grains as old as 4.4 Ga have been found [14–16]. The outcrop in which these detrital zircons originated has been eroded and is missing

from the record, and only sediments remain. These zircons are gaining global attention for the innovative research undertaken at the Jack Hills to assess the conditions of Hadean Earth and the beginnings of biotic activity. Such highly stable zircon minerals, formed by fractionation of a hydrated crust, are generally common in granitic but rare in mafic rocks. The detrital zircon grains from the Jack Hills suggest that a continental granitic crust existed during the Hadean Eon (~4.4 Ga). Similarly, both detrital zircons [17, 18] and zircon xenocrysts trapped inside a granitic rock [18] from the Singhbhum Craton of India indicate the presence of a Hadean (~4.2 Ga) source rock whose formation required water near Earth's surface. Oxygen isotopes in the zircons provide valuable information about the existence of liquid water interacting with the granitic crust 4.3 billion years ago.

The zircon crystals found in the Singhbhum Craton and the Jack Hills are significant for their extreme age. However, they have also provided geochemical information that has extended our understanding of the geological processes that occurred during Earth's initial formation. By analyzing the oxygen isotopic ratios (oxygen-18/oxygen-16) in detrital zircons from the quartzitic rocks in the Murchison district of Western Australia, some have suggested that Earth contained liquid water in some form as early as 4.3 Ga [14]. The same detrital zircons support the existence of an evolved granitic crust on Earth as soon as 4.4 Ga, but the paucity of Hadean zircons suggests that zircon-bearing felsic rocks, such as the tonalite–trondhjemite–granodiorite (TTG) rocks that make up the majority of Archean crust, may not have been a prominent component of Earth's earliest crust. Perhaps a few protocontinents or island continents with felsic crusts began to emerge in the global ocean and were subjected to erosion, forming sedimentary deposits around 4.4–4.2 Ga, as interpreted from the detrital zircons. Hadean Earth is characterized by a thick basaltic crust covered by an ocean, with very little in the way of drylands, mainly composed of felsic rocks. The picture emerging from the detrital zircons suggests that Hadean Earth, although battered with impacts, became a watery world with normal recycling processes and thus a more habitable environment than previously believed [14]. Nevertheless, during the Late Heavy Bombardment period, the primordial ocean would have been occasionally vaporized locally by massive meteorite impacts.

5.6 Collision and Creation: The Building Blocks of Life

Earth's status as the only life-sustaining planet in the solar system is the result of the timing and delivery mechanism of the building blocks of life to the silicate crust of the young planet. Many terrestrial formation models suggest that Earth on its own would have formed extremely poor volatiles and was deficient in the organic compounds necessary for life synthesis. Surprisingly, the chemical makeup of life is closer to the chemistry of meteorites and stardust than it is to that of our rocky planet during its early history. Comets and carbonaceous asteroids are rich in organic molecules, facilitating the emergence of life on early Earth. Recent research in space exploration and astrobiology has provided strong evidence that meteorite impacts strongly facilitated the emergence of life on early Earth. Life's biochemical components can form in the harsh, zero-gravity environment of deep space, bolstering the odds that massive meteorite strikes on young Earth might have helped trigger life. Unlike conditions in space, early Earth's high gravity and hot temperature were not conducive to creating these compounds on the planetary surface. Therefore, it is likely that a significant fraction of the surface inventory of organic material was brought down to Earth through such bombardment during the Hadean–Archean transition. Interstellar molecules provided the primary organic material deposited on all solar system bodies and delivered to the surface of Earth.

To understand what chemical compounds would have been available prior to the origin of life, we turn our attention to carbonaceous chondrites and comets. As we have noted, complex biomolecules such as lipid membranes, amino acids, nucleobases, phosphorous, and sugars have been detected in carbonaceous meteorites [19]. Moreover, these compounds are abundant in interstellar space. Complex organic molecules, precursors to life, have been detected everywhere in space—in comets, carbonaceous asteroids, and interstellar dust. The Stardust mission detected these building blocks in interstellar dust particles [20]. In interstellar space, far from the Sun, temperatures are freezing, leading to abundant ice and other frozen gases such as methane, ammonia, carbon dioxide, carbon monoxide, and even ethyl alcohol. Throughout the Heavy Bombardment period, this volatile arrived on Earth from distant reservoirs in the outer solar system.

New experiments simulating conditions of deep space reveal that the complex prebiotic compounds could have been created in icy interplanetary dust in zero gravity and eventually carried to Earth. These experiments suggest that the building blocks of life may have begun in a rather unusual freezing environment of interstellar space during the planetary formation and may have been delivered to early Earth in the suitable settings provided by meteorites. At the National Aeronautics and Space Administration (NASA) Ames Research Center, interstellar ice simulation experiments in a cryogenic laboratory at 40 K (−233 °C) have shown that ultraviolet (UV) radiation processing of pre-solar ices—interstellar materials formed prior to the birth of the Sun—leads to more complex organic compounds. In these experiments, all the primary building blocks of life, including ribose and related sugars, were synthesized via irradiation of interstellar ice analogs (silicate minerals coated with

simple molecules such as H_2O, CO, CO_2, NH_3, and CH_3OH) [4, 21, 22]. Ribose is the central molecular subunit of RNA, so its prebiotic origin in interstellar environments is significant. These experimental findings support the identification of organic molecules in comet samples collected in situ by the 'Philae lander' part of the cometary Rosetta mission [23]. In another laboratory simulation of interstellar ices, the building blocks of comets, a frozen mixture of water, methanol, ammonia, and carbon monoxide, were subjected to UV radiation [24]. This combination yielded organic compounds that, when immersed in water, formed bubbles reminiscent of the cell membranes that enclose and concentrate the chemistry of life. Many of the organic compounds included in these experiments, some of which are necessary to the origin of life, are also present in meteorites and cometary and asteroidal dust.

In space, when atoms are locked in ice, bond-breaking processes can make molecular fragments recombine into unusually complex structures in ways that would not be possible if these fragments were free to drift apart. Everywhere in space, complex organic compounds form on these ice grains, especially in UV-rich regions around young suns. Much of the organic compounds found on early Earth had an interstellar heritage. Carbon-based compounds are synthesized everywhere in the Milky Way, including diffuse interstellar clouds, giant molecular clouds, and the protoplanetary disks forming new solar systems. Meteorites and interstellar dust provide a record of the chemical processes that occurred before life began on Earth. Like life rocks from on high, triggering, and accelerating biosynthesis on Earth, meteorites are otherworldly, providing a physical connection between heaven and Earth. Most of the biosphere's organic materials—like our own bodies—arrived as stardust, a gift from the meteorites.

We have been tracking the recent paradigm shift in origin-of-life research, from the Miller–Urey type of experiment simulating hypothetical conditions of Earth's early atmosphere to demonstrate that organic compounds could have formed spontaneously on young Earth for the study of cosmic building blocks, delivered by meteorites, as the likely source of life's ingredients. Experiments by NASA suggest that the building blocks of life originated in a cold interstellar medium of extremely low density. These differences in temperatures and media may explain why the Miller-type experiments in the laboratory failed to produce similarly diverse arrays of organic compounds.

We believe that life's building blocks arrived ready-made from space. For example, cosmic dust or micrometeorites (<2 mm) would have delivered some of these biomolecules to early Earth. Some of the organic compounds they contained—such as polycyclic aromatic hydrocarbons (PAHs), aliphatic chains, and amino acids—were used in life synthesis. They represent the enormous mass flux of extraterrestrial material (30,000 tons per year), raining down on Earth daily, carpeting the land and water surface with fine dust [25]. As another example, about 5% of chondrites or rocky asteroids are carbonaceous chondrites, which are especially relevant to the origin of life. Carbonaceous chondrites, comets, and micrometeorites in interstellar clouds are extremely rich in prebiotic compounds and volatile elements like carbon and nitrogen that are essential to organic life.

Among the most celebrated carbonaceous chondrites is the Murchison meteorite that fell to Earth near Murchison, Victoria, in Australia in 1969, only two months after the Apollo 11 splashdown (Fig. 5.4). It was an unexpected heavenly gift to NASA's cosmochemists at Ames Research Center in California, where the Murchison meteorite was analyzed by a team headed by Dr. Cyril Ponnamperuma. The results were of the greatest scientific importance. More than 90% of the organic material is a kerogen-like polymer composed of polycyclic aromatic hydrocarbons (PAHs), some derivatives involved in life processes. About 0.5% of the Murchison meteorite contains all the biologically relevant compounds—those that are soluble in water and organic solvents. Materials detected in the Murchison meteorite that are essential components of living cells include pyrimidine and purine nucleobases, sugars, and phosphates for making nucleic acids, 70 amino acids, 6 of which are components of proteins and membrane-forming compounds such as lipids, long-chain monocarboxylic acids, and carbohydrates, along with a small amount adenine, one of the bases of nucleic acids [6, 21]. If these organic compounds were prevalent in the parent asteroid from which meteorites were derived, it makes sense that similar sets of organic compounds were delivered directly to early Earth during the Heavy Bombardment period.

These asteroids contained liquid water for the first 10,000 years of their history, during which the synthesis of their amino acids started. Yet, there is no evidence of peptides or nucleic acids in these asteroids, and, in general, it appears that prebiotic synthesis in these objects did not proceed beyond monomers. However, starting with these cosmic monomers, prebiotic chemistry then continued much further on Earth itself, first leading to polymers and then eventually to the first cells. These cosmic monomers must have survived their catastrophic impacts during the collision with Earth. To this day, carbonaceous chondrites collected across the planet still contain those essential building blocks, clearly demonstrating that the hitchhiking organic molecules withstood the extreme pressures and temperatures of fiery crashes onto Earth's surface, remaining intact despite tremendous shock waves and other violent conditions [26]. In addition to carbonaceous chondrites, iron-containing asteroids were a rich source of iron for initial metabolism. Impacts mobilize iron minerals, particularly iron sulfides, with potential consequences for prebiotic reactions in the iron–sulfur world.

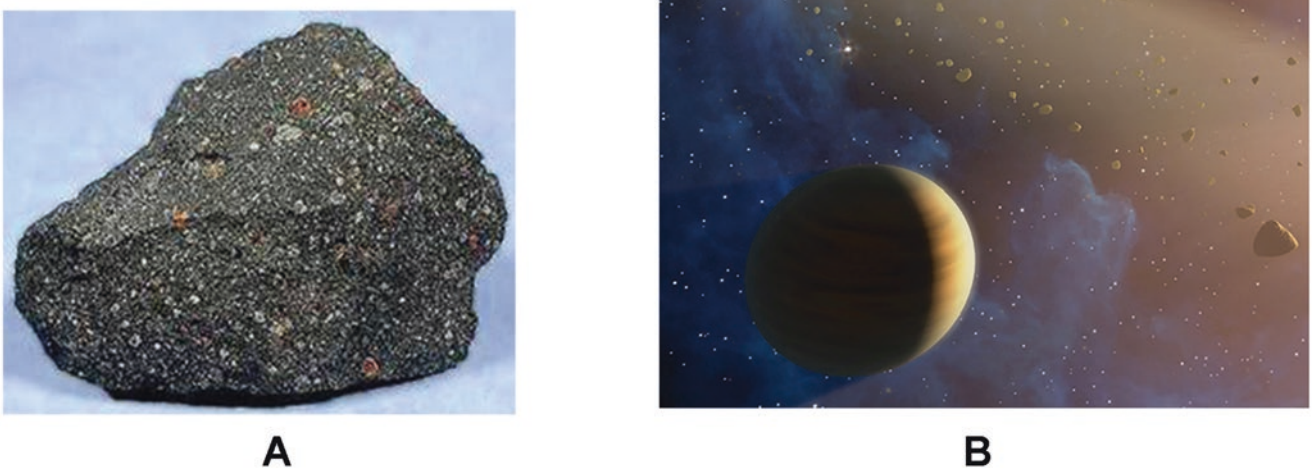


Fig. 5.4 (**a**) The Murchison meteorite (carbonaceous chondrite) is rich in organic compounds containing more than 70 kinds of amino acids and a smaller amount of amphiphilic material for the lipid membrane, even smaller amounts of nucleobases such as purines and pyrimidines, and simple carbohydrates such as ribose. (**b**) Asteroids battered young Earth during a violent period of the Late Heavy Bombardment around 4.1–3.9 billion years ago. (Courtesy NASA)

We have noted that comets differ from asteroids in being 'dirty snowballs' composed of ice, CO_2, and silicate mineral dust, that occur in the outer solar system. However, like asteroids, they are leftovers from the dawn of the solar system. Comets contain 20% organic material; much of it is a black polymer-like kerogen. The chemical composition of comets provides vital information about the raw materials of organic compounds of prebiotic interest present in interstellar space, including amino acids, adenine, ketones, quinones, carboxylic acids, hydrogen cyanide (HCN), polycyclic amino aromatic hydrocarbons, thioformaldehyde, acetaldehyde, sugars, cyanogen, cyanide, methanol, and ethanol [27]. HCN is a critical molecule in the prebiotic synthesis of amino acids, nucleobases, and sugars. Numerous complex organic molecules, such as ethyl alcohol and sugar, have been detected in the comet C/2014 Q2 (Lovejoy) [28]. Recently, the 'Rosetta mission' of the European Space Agency (ESA) craft Philae lander has detected water and organic molecules on the surface of 67P/Churymov–Gerasimenko, a 4-km-wide comet [29]. In all, 16 organic compounds were identified, divided into 6 classes of organic molecules, namely, alcohols, carbonyls, amines, nitriles, amides, and isocyanates. The presence of complex organic compounds in comets and asteroids implies that the early solar system fostered prebiotic material in noticeable concentrations. Meteorites and stardust offer natural examples of purely prebiotic chemistry occurring in space.

Significant impacts created tremendous shock waves and other violent conditions, causing volcanic activity in a water-rich geological system in early Earth environments and vaporizing oceans as they hit [30, 31]. Those bolides that survived the impact brought prebiotic molecules with them. Comets would have been ideal packages for delivering terrestrial volatiles, including some critical precursors for prebiotic synthesis. Because of the icy cushion of comets, the building blocks of life would have remained intact upon impact, whereas others would have fused with other chemicals to form more complex organic compounds [32]. We believe that organic compounds enjoyed multiple delivery systems—comets, asteroids, and the fall of interstellar dust—over the millions of years of the Eoarchean period. Mixing these cosmic ingredients with various biomolecules produced in the hydrothermal crater basin created the primordial soup [4, 33]. The initial stages in the origin of life required a continuous source of organic molecules, both cosmic and terrestrial, to act as precursors to the biopolymers [6]. A constant rain of interplanetary dust particles continues to this day, delivering the same organic compounds that contributed to the formation of the primordial soup on early Earth [30, 32].

Life as we know it requires three essential ingredients: water, organic matter, and energy. A carbonaceous chondrite carrying organic material that crashes into Earth could supply all these ingredients in a tidy package. Could these organic molecules in meteorites survive the catastrophic impacts during the collision with Earth? Surprisingly, they did withstand high-speed impacts as organic-rich carbonaceous chondrites such as the Murchison meteorite suggests [5, 6, 26]. Biomass preservation in impact melt ejecta supports this contention [33]. The degree of organic material

preserved in this impact melt with high fidelity is surprising. In summary, meteorites played a critical role in the delivery of the building blocks of life because organic compounds were able to survive the high pressure and temperature of impact events. Comets would have been ideal vehicles for the delivery of terrestrial volatiles, including critical precursor molecules for prebiotic synthesis.

5.7 The Solvent of Life

The story of liquid water is an indispensable chapter in the story of life. Without liquid water, life as we know it simply would not exist. Oceans envelop most of Earth's surface and drive the water cycles that modulate our lands and atmosphere. A bacterial cell is 70% water by weight. As it is capable of transporting and mixing organic molecules, liquid water is a universal solvent in the delicate chemistry that makes life possible. Water—H_2O—is a small, highly polar molecule capable of dissolving many kinds of molecules, preserving their integrity while transporting them to reaction sites. Ice or water vapor cannot accomplish the same function. Water in the form of ice is relatively common everywhere in the universe, from vast interstellar clouds to many planets and moons. In contrast, liquid water is rare in the cosmos. Thus, the origin of life greatly hinges on the source and availability of liquid water on early Earth. Water on Earth is kept in its liquid form by the precise combination of atmospheric pressure and internal heat.

From geological records, it appears that Earth became a watery planet about 4 billion years ago, or even earlier, but debates are still ongoing about where all that life-sustaining liquid came from [14, 19]. Astronomers know that interstellar water is abundantly available to young planetary systems, and our young Earth collected plenty of it. However, despite the dominance of water on Earth, there is a great deal of controversy about its origin on our planet. We do not know whether the planet's water was present at the time of Earth's formation, formed through the degassing of volcanoes, or arrived later, delivered by asteroids or comets. It seems most likely that Earth's water did not originate on our planet. Over billions of years, comets and asteroids have collided with Earth, enriching our world with water and organic compounds.

If meteorites did bring the oceans and the atmosphere to Earth, then we might expect some chemical signature of that contribution to the planetary surface. In fact, the composition of seas offers some clues to their origin. Mounting evidence suggests that the planet's water arrived aboard meteorites around the beginning of the Late Hadean Bombardment. Our seawater may well contain some trace of that ancient event. Some researchers, such as Armand Delsemme, championed the view that an intense bombardment of comets brought water to primitive Earth [27]. However, chemical markers in ocean water indicate that most of it came from asteroids. Water vapor streaming off comets contains a higher percentage of deuterium (heavy hydrogen, ^2H) than does the water on Earth. A new analysis of the isotopic ratio of hydrogen of the carbonaceous chondrites suggests that water did not originate from the comets in the outer solar system but from the asteroid belts between Mars and Jupiter. The deuterium-to-hydrogen (D/H) ratio in ocean water, approximately 150 ppm, is like the average carbonaceous chondrites, suggesting that these bodies are the principal sources of water, not icy comets [34]. A recent study of the D/H ratio of the comet 67P/Churymov–Gerasimenko has supported this view of water's asteroidal origin [35]. Most water-rich asteroids are carbonaceous chondrites, which can contain up to 28% water. Carbonaceous asteroids are thus likely to be the main sources of Earth's volatiles and the preferred vehicle for the delivery of water to the early planet.

The asteroid belt is generally considered as the domain of rocky bodies, being too close to the Sun for ice to survive. However, astronomers using NASA's Infrared Telescope Facility have recently detected a coating of ice and complex organic compounds on one of the largest asteroids, called 24 Themis. This discovery endorses the idea that asteroids may have brought plentiful supplies of water and organic material to Earth during the Heavy Bombardment period [36]. It thus appears that the asteroidal bombardment of early Earth delivered not only the building blocks of life but also water and other essential chemicals such as formaldehyde, hydrogen cyanide, and ammonia. The presence of the vital ingredients for life on ancient carbonaceous chondrites, about 4.5 billion years old, supports this view. The abundance of the deuterium isotope in hydrated minerals found in certain meteorites is identical to the deuterium content of ocean water, suggesting that the outer main belt of asteroids is the most likely source of that water [34, 37]. The millions of carbonaceous asteroids that crashed into Earth during the Late Heavy Bombardment could have brought the organic compounds and water to make our planet more habitable by creating oceans, innumerable crater lakes, and a wet rather than the arid atmosphere. Abundant water facilitated the prebiotic reaction medium on early Earth. The delivery of cosmic ingredients led to a cascade of reactions that started the prebiotic syntheses in the cradles of life.

All life, including our own, has been interconnected since the birth of the first cells with the vast solar system we all inhabit. Most of the material of which we are made comes from the debris of dying stars. The supernova explosion showered our planet with ready-made cosmic ingredients for some form of carbon-based life to develop, from simple bacteria to intelligent observers. The very elements of which we are made preserve our cosmic connections—those atoms formed in stars over billions of years and multiple star lifetimes. Moreover, although we really are made of stardust, that is not the entire story. Recent cosmic exploration has

suggested that some of the hydrogens that make up 9.5% of our bodies, and some of the lithium that our bodies contain in tiny trace amounts, originated from the Big Bang. The elements of life can be as old as the universe itself. As the poet Walt Whitman grasped during the golden age of American astronomy—inspired by Robert Chamber's treatment of the nebular hypothesis, *Vestiges of the Natural History of Creation*, anonymously published in 1844—everything in the cosmos is interconnected. In his immortal poem *Song of Myself*, Whitman proclaimed a faith that contemporary science has confirmed: 'I believe a leaf of grass is no less than the journey-work of the stars.'

5.8 Conclusions

Astronomers speculate that a shock wave from a catastrophic supernova explosion about 4.6 billion years ago may have sparked the birth of our solar system—its planets, moons, asteroids, and comets—when it crashed into a swirling cloud of interstellar gas and dust. The solar system is believed to have coalesced from this immense rotating dense cloud of gas and dust. Residues of matter ejected from the ancient blast can still be found today in samples of interstellar dust, primitive asteroids such as carbonaceous chondrites, and comets. They also contain the atomic and molecular seeds of life. Short-lived radioactive isotopes from the early solar system found in asteroids support this supernova explosion scenario. Asteroids are rocky debris remnants from the era of planet formation; they contain materials from which Earth and the other planets were initially made, frozen in time. They include life's secrets. The oldest material inside these asteroids is 4.5 billion years old, which is how astronomers can estimate how long ago the nebula that formed the solar system started to coalesce. Asteroids still orbit the Sun in a tightly packed belt between Mars and Jupiter. Comets are in the Oort Cloud that rings the outer fringe of the solar system. They are balls of ice and rock that grow tails as they approach the Sun during their highly elliptical orbits. Both carbonaceous chondrites and comets are rich in organic compounds, the building blocks of life. The presence of complex organic compounds in asteroids and comets implies that prebiotic material is present in meteorites and interstellar dust, a gift from that supernova explosion.

The earliest phase of Earth's evolution occurred during the 500 million years of the Hadean Eon. Earth was then a molten globe enduring the continuous bombardment of meteorites and the release of heat from radioactive decay. During this time, Earth's core, mantle, and earliest crust began to differentiate into three layers by gravitational settling. The core formed as iron and nickel sank through the lighter silicate minerals of the crust to the center of the planet, creating a magnetic field around our planet. Stretching from the planet's core out into space, this magnetic field shields Earth from the deadly high-energy solar winds emitted by the Sun. The Moon also originated during the Hadean period, most likely due to a collision between Earth and a planet named Theia that perished from that impact. No Hadean rocks survive other than some detrital zircons as old as 4.4 Ga that provide tantalizing evidence of the primordial presence of granitic crust and liquid water.

The building blocks of life had their beginnings in the stardust in interstellar space during the explosion of a supernova and brought down to Earth by meteorites. Some of the building blocks have been detected in the Murchison asteroid and interstellar dust, including lipids, amino acids, nucleobases, sugars, phosphates, hydrocarbons, and carbonyl compounds—the most basic and important components of life. Comets contain prodigious quantities of organic compounds. Life appears to have originated at the end of the most violent geological period of Earth's history, the Late Heavy Bombardment period (~4.1–3.8 Ga), when asteroids battered Earth about 1000 times more frequently than they do today, delivering critical organic molecules and water to set probiotic chemistries going.

Our connection with the cosmos is preserved in our biochemistry. The nucleosynthesis of two primordial elements—hydrogen and helium—took place during the origin of the universe. The other heavier elements appeared only later with the nuclear fusion and collapse of the first generations of stars. During a supernova explosion, a dying star produces six prebiotic elements present in all life-forms—carbon, hydrogen, oxygen, nitrogen, phosphorous, and sulfur—and then disperses them into the interstellar medium. While the hydrogen of our world may have come from the Big Bang event about 14 billion years ago, meteorites and interstellar dust delivered most of these primary elements, along with the organic compounds and the water needed for life synthesis, directly to Earth. The ties that bind the cosmos and life on Earth cannot be loosed. Our bodies are stardust as old as the universe but enriched during the time of a supernova explosion about about 4.6 billion years ago.

References

1. Banerjee P, Qian YZ, Heger A, Haxton WC (2016) Evidence from stable isotopes and ^{10}Be for Solar System formation triggered by a low-mass supernova. Nat Commun 7:13639. https://doi.org/10.1038/ncomms13639
2. Boss AP, Keiser SA (2010) Who pulled the trigger: a supernova or an asymmetric giant branch star? Astrophys J Lett 717:L1–L5
3. Brennecka GA, Borg LE, Wadhwa M (2013) Evidence for supernova injection in the solar nebula and the decoupling of *r*-process nucleosynthesis. Proc Natl Acad Sci U S A 110:17241–17246
4. Chyba C, Sagan C (1992) Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origin of life. Nature 355:125–132

5. Pizzarello S, Cronin JR (2000) Non-racemic amino acids in the Murchison and Murray meteorites. Geochem Cosmochem Acta 64:329–338
6. Deamer DW (2011) First life: discovering the connections between stars, cells, and how life began. University of California Press, Berkeley
7. Sandford SA, Bera PB, Lee TJ, Materese K, Nuevo M (2015) Photosynthesis and photo-stability of nucleic acids in prebiotic extraterrestrial environments. Top Curr Chem. https://doi.org/10.1007/128_2013_499
8. Bischoff A, Geiger T (1995) Meteorites for the Sahara: find locations, shock classification, degree of weathering and pairing. Meteoritics 30:113–122
9. Kring DA, Cohen BA (2002) Cataclysmic bombardment throughout the inner Solar System. J Geophys Res 107. https://doi.org/10.1024/2002JE001529
10. Marchi S, Bottke WF, Elkins-Tanton LT, Bierhaus M, Wuennemann K, Morbidelli A, Kring DA (2014) Widespread mixing and burial of Earth's Hadean crust by asteroid impacts. Nature 511:578–582
11. Cohen BA, Swindle TD, Kring DA (2000) Support for the lunar cataclysm hypothesis from lunar meteorite impact melt ages. Science 290:1754–1756
12. Halliday AN (2006) The age of the Earth – what's new? Elements 2:205–210
13. Frankel C (1996) Volcanoes in the Solar System. Cambridge University Press, Cambridge
14. Mojzsis SJ, Harrison TM, Pidgeon TT (2001) Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. Nature 409:178–18115
15. Wilde SA, Valley JW, Peck WH, Graham CM (2001) Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. Nature 409:175–178
16. Harrison TM, Schmitt AK, McCulloch MT, Lovera OM (2008) Early (> 4.5 Ga) formation of terrestrial crust: Lu-Hf, $\partial^{18}O$, and Ti thermometry results for Hadean zircons. Earth Planet Sci Lett 268:476–486
17. Miller SR, Mueller PA, Meert JG, Kamenov GD, Pivarunas AF, Sinha AK, Pandit MK (2018) Detrital zircons reveal evidence of Hadean crust in the Singhbhum Craton, India. J Geol 126:541–552
18. Chaudhuri T, Wan Y, Mazumder R, Ma M, Liu D (2018) Evidence of enriched, Hadean mantle reservoir from 4.2-4.0 Ga zircon xenocrysts from Paleoarchean TTGs of the Singhbhum Craton, eastern India. Science Rep 8:7069. https://doi.org/10.1038/s41598-018-25494-6
19. Bernstein MP, Sandford SA, Allamonda LJ (1999) Life's far flung raw material. Sci Am 263:42–49
20. Brownlee DE, Tsou P, Anderson JD, Hanner MS, Newburn RL, Sekannia Z, Clark BE, Horz F, Zolensky ME, Kissel J, McDonnell JAM, Sandford SA, Tuzzolino AJ (2003) Stardust: comet and interstellar dust sample return mission. J Geophys Res 108(E10):8111. https://doi.org/10.1029/2003/E0002087
21. Deamer DW, Dworkin JP, Sandford SA, Bernstein MP, Allamandola LJ (2002) The first cell membranes. Astrobiology 2:371–381
22. Meinert C, Myrgorodska I, Mercellus PD, Buhse T, Nahon L, Hoffmann SV, d'Hendercourt LLS, Meierhenrich UJ (2016) Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. Science 352:208–212
23. Altwegg K, Balsiger H, Ber-Nun A, others (2016) Prebiotic chemicals—amino acid and phosphorous—in the coma of comet 67P/Churymov-Gerasimenko. Science 2. https://doi.org/10.1126/sciadv.1600285
24. Dworkin JP, Deamer DW, Sandford SA, Allamonda LJ (2001) Self-assembling amphiphilic molecules: synthesis in simulated interstellar/precometary ices. Proc Natl Acad Sci U S A 98:815–819
25. Motrajit G, Guan Y, Leshin L, Taylor S, Genge M, Joswiak D, Brownlee D (2006) Oxygen isotope measurements of bulk unmelted Antarctic micrometeorites. Geochim Cosmochim Acta 70:4007–4018
26. Callahan MP, Smith KE, Cleaves SH, Ruzicka J, Stern JC, Glavid DP, Huse CH, Dworkin JP (2011) Carbonaceous meteorites contain a wide range of nucleobases. Proc Natl Acad Sci USA 108:13995–13998
27. Delsemme AH (2001) Argument for the cometary origin of the Biosphere. Sci Am 89(9):431–442
28. Biver N, Bocklee-Morvan D, Moreno R, Crovisier J, Colom P, Lis DC, Sandquist A, Boissier J, Despois D, Milan SN (2015) Ethyl alcohol and sugar in comet C/2014 Q2 (Lovejoy). Sci Adv 1:e1500863
29. Goesmann F, Rosenbauer H, Bredhoft JH, Cabane M, Eherenfreund P, Gautier T, Giri C, Kruger H, Roy LL, McKenna-Lawlor S, Melenrheinrich UJ, Caro GMM, Raulin F, Roll R, Steele A, Steininger H, Stenberg R, Szopa C, Thiemann W, Ulamec S (2015) CHO-bearing organic compounds on comet 67P/Churyumov-Gerasimenko reveale by COSAC mass spectrometry. Science:349. https://doi.org/10.1126/science.aab0689
30. Sleep NH, Zahnle KJ, Kasting JF, Morowitz HJ (1989) Annihilation of ecosystems by large asteroid impacts on the early Earth. Nature 342:139–142
31. Zahnle K, Arndt N, Cockell C, Halliday A, Nisbet E, Selsis F, Sleep NH (2007) Emergence of a habitable planet. Space Sci Rev 129:35–78
32. Damer B, Deamer DW (2015) Coupled phases and combinatorial selection in fluctuating hydrothermal pools: a scenario to guide experimental approaches to the origin of cellular life. Life 5:872–887
33. Howard KT, Bailey MJ, Berhanu D, Bland PA, Cressey G, Howard LE, Jeynes C, Matthewman Z, Sephton MA, Stolojan V, Verchvsky S (2013) Biomass preservation in impact melt ejecta. Nat Geosci 6. https://doi.org/10.1038/NGEO1996
34. CMO'D A, Bowden R, Fogel M, Howard KT, CDK H, Nittler LR (2012) The provenances of asteroids, and their contributions to the volatile inventories of the terrestrial planets. Science 337:721–723
35. Altwegg K et al (2014) 67P/Churyumov-Gerasimenko, a Jupiter family comet with a high D/H ratio. Sci Express. https://doi.org/10.1126/Science.1261952
36. Rivkin AS, Emery JP (2010) Detection of ice and organics on an asteroidal surface. Nature 464:1322–1323
37. Sarafian AR, Nielson SG, Marschall HR, McCubbin FM, Monteleone BD (2014) Early accretion of water in the inner Solar System from a carbonaceous chondrite-like source. Science 346:623–626

The Cradle of Life

6

I cannot remember my mother
Only sometimes amid my play
A tune seems to hover over my playthings
The tune of some song that she used to
Hum while rocking my cradle.

—Rabindranath Tagore, 1922

6.1 Geological Constraints on the Origin of Life

Where did life begin? The key to discovering how life originated is to first establish where on this planet that happened in the prebiotic environment four billion years ago. Researchers studying the origin of life each have their favorite geological site of life's cradle. Each setting has its pros and cons. Life could have started in a range of environments, from the bottom of the oceans and hot springs on volcanic islands to hydrothermal crater lakes on protocontinents. However, is it possible that one likely special place had the perfect combination of all the conditions essential for life to get started?

Various environments for the origin of life during the Late Heavy Bombardment have been proposed, such as the anaerobic atmosphere, a global ocean, hydrothermal systems, photochemistry, and high-energy radiation during the Eoarchean. The primitive crust had abundant cosmic ingredients delivered by meteorites and interstellar dust to orchestrate prebiotic synthesis. The components of life would have arrived from space on early Earth practically preassembled and freeze-dried, but these cosmic ingredients needed the right crucible for their chemical evolution. Life on Earth may have had several false starts. With the widespread availability of cosmic molecules, the synthesis of life could have started simultaneously in different environments but then failed to progress very far, except for that one spot where life established a secure foothold. We will consider all the possible sites on early Earth where abiogenesis might have started, eliminate those that conflict with the available evidence, and select the most plausible candidate. Once the bombardments ceased, it would have been possible for life to proliferate. Nevertheless, undoubtedly, the environment of young Earth was severe and inhospitable, so life's emergence must have been confined to an extremely sheltered location capable of protecting fragile organic molecules from natural hazards.

Our planet is estimated to be 4.55 billion years old, and the earliest known microbial fossils formed nearly four billion years ago. We seek to understand the origin of life in a narrow window of several million years. Which environment of young Earth was best suited to promote prebiotic synthesis and complexity? The environment encompasses the composition of the atmosphere, the acidity and salinity of primal oceans, and the existence of drylands such as protocontinents and volcanic islands. I will begin with the geology of Eoarchean Earth. We believe that the first life appeared during the beginning of the Eoarchean era or even earlier. At that time, the primordial ocean dominated Earth's landscape, punctuated by occasional islands jutting above the water. However, there is still no professional consensus on the precise environment that fostered the beginning of life. Debates are ongoing between geologists and chemists whether life began on land or at sea.

To understand the geological site of life's beginnings, we must reconstruct Eoarchean Earth, where a permanent crust capable of fostering the first life began to form. We can begin by building up a picture of Archean Earth and its geological, atmospheric, and biological conditions. Gases released by impacts and eruptions formed a thick atmosphere mainly consisting of carbon dioxide (CO_2), with small amounts of nitrogen, methane (CH_4), water vapor, and sulfur gases but no oxygen to speak of. The concentrations of greenhouse gases such as carbon dioxide and methane were sufficient to offset a fainter Sun by trapping more of its heat in the atmosphere. With a hundred times more methane than there is in the atmosphere today, the Archean sky would have had a faintly pinkish tinge (see book's cover). When the first life emerged, methanogenic bacteria could readily have maintained this high methane concentration.

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_6

Archean Earth was anoxic, with high heat flow from the mantle. There was no free oxygen in ocean water or atmosphere. Its oceans would have been mildly acidic, warm, and extremely salty. The analysis of zircons, durable minerals of the primordial continental crust, suggests that while continents and oceans began to form in the Hadean Eon [1], by the time of the Eoarchean, Earth's surface was cool enough to have a reliably permanent crust and liquid water to form stable oceans. We know from crater impacts and lunar samples that the Moon and Earth both suffered heavy bombardments during the beginning of the Eoarchean. Still, since the scars of these impact events have long been obliterated, the impact record on Earth remains obscure. The Eoarchean record is fragmentary because those catastrophic events and later recycling by plate tectonics must have destroyed almost all of early Earth's original structure.

In geology, the older the rock, the lesser the chance of its preservation. Phanerozoic rocks surviving to the present are abundant—those of the Proterozoic, less so and those of the Archean, rare. Those as old as Eoarchean, bearing the earliest known fossils, are nearly nonexistent. The rarity of an ancient crust older than 3.6 Ga makes it challenging to describe the crust-forming processes that operated during Earth's earliest history. No crustal rocks survive from the time of the Heavy Bombardment. The oldest known crust is the Acasta Gneiss in northwestern Canada (~4.03 Ga) [7]. However, while the available evidence is sketchy, in recent years, several scientific research fields have brought forward support for the origin of life not, as previously widely favored, in deep oceans, but in a terrestrial setting, on one of the protocontinents.

6.2 A Habitable Eoarchean World

The intense impact environment of Hadean (~4.5–4 Ga) Earth, an eon of extreme geological violence, determined in large part the initial physical and chemical states of the planet. For much of the Hadean, cosmic impacts pummeled Earth and its sister worlds in the inner solar system. During the Late Hadean, the basaltic crust began to form from the magma ocean and was then covered by the global sea [1]. The nature of the geodynamic environment in which Earth's first continental granitic crust formed and stabilized remains controversial. Evidence of the Hadean crust is provided by detrital zircons from Australia and India, the oldest surviving primordial crustal material. Grains as old as 4.4–4.2 Ga have been discovered. The granitic outcrop in which the detrital zircons originated has been destroyed, and only sediments remain [2, 3]. Overall, late Hadean Earth most probably consisted of a thick basaltic crust covered by a global ocean, punctuated by few island continents, mostly of granitic rocks—the source of detrital zircons.

If liquid water and elevated crust existed as early in Earth's history as in the late Hadean Eon, was life already present then? It is impossible to confirm whether life arrived before the Archean period because the Late Heavy Bombardment (LHB), which cratered our Moon, obliterated rocks older than the Archean eon. The oldest known fossils come from rocks nearly four billion years old. There is no question that Earth was an extremely different place during the Eoarchean. There were little emergent drylands, except for several island continents or protocontinents, highly battered by meteorite impacts. Nevertheless, we believe that life began around this time. The first life-forms, reasonably believed to be hyperthermophiles that are organisms that can tolerate temperatures up to 120 °C, were suited to survive the LHB by colonizing protected hydrothermal vent habitats. As we gain more knowledge of what it takes to make a planet habitable, we can also apply our understanding of the conditions on Earth when life first arose to our hunt for other life in our solar system and beyond.

6.2.1 Eoarchean Crustal Evolution

How did the continental crust on Eoarchean Earth arise? Due to fragmentary Eoarchean and virtually absent Hadean rock records, the nature of the continental crust is difficult to determine. Moreover, we do not know the major tectonic forces operating at that time. Following the Hadean, the second stage in Earth's geological evolution occurred during the Eoarchean Eon in a pre-plate tectonic regime [4]. Elevated mantle temperatures compared to the present day weakened the lithosphere by the emplacement of melts that inhibited subduction and hence plate tectonics. Impact-induced mantle upwellings that percolated the lithosphere drove vigorous magmatism. During this period, the crust was at least 15–20 km thick [5]. A high geothermal gradient in the early Archean led to crustal buoyancy, increased crustal ductility, and reduced crustal strength. Even while a relatively thin, ductile, and buoyant Archean lithosphere would have prevented subduction, a vertical mantle force unrelated to plate tectonics may have cycled material from the mantle to the surface.

Although ancient zircons from the Hadean suggest that Earth's continental crust existed as long ago as 4.4 Ga, the oldest rock formations exposed on Earth's surface are exclusively Archean. The Eoarchean Eon witnessed the formation of the early continental felsic crust and global oceans, the first earthly ecosystems, the emergence of life, and fundamental changes to the atmosphere. During the early Eoarchean, the meteoritic pummeling was not over, but as the frequency of impacts slowed down, Earth became more tranquil. The planet cooled, clouds formed, torrential rains submerged the lowlands to form global oceans, and the crust began to harden, creating 'terra firma,' followed by erosion

and sediment recycling. While Earth cooled, the planet's first outer layer, the lithosphere, was a single, stable, deformable shell that lacked plate tectonic activity. It was a one-plate world devoid of plate tectonics [4].

The primordial oceans surrounding the emerging protocontinents covered 90% of Earth's surface. Water-lain sediments and zircons on the ancient Archean crust indicate that liquid water was prevalent as early as four billion years ago or even earlier [1, 6]. Earth was no longer an inhospitable hellscape; it was gradually turning into a viable environment. A desolate orb was transforming into a tranquil abode. As endless rains submerged the lands, the new oceans gathered organic molecules from an atmosphere largely composed of carbon dioxide and nitrogen. Impacting bodies and the degassing of the mantle made the first oceans highly saline. The world was watery from pole to pole, with only a few sparsely distributed continental islands formed by impact melting and plume volcanism. Filling much of Earth's night sky, the Moon cast a startling shadow on the planet (Fig. 6.1).

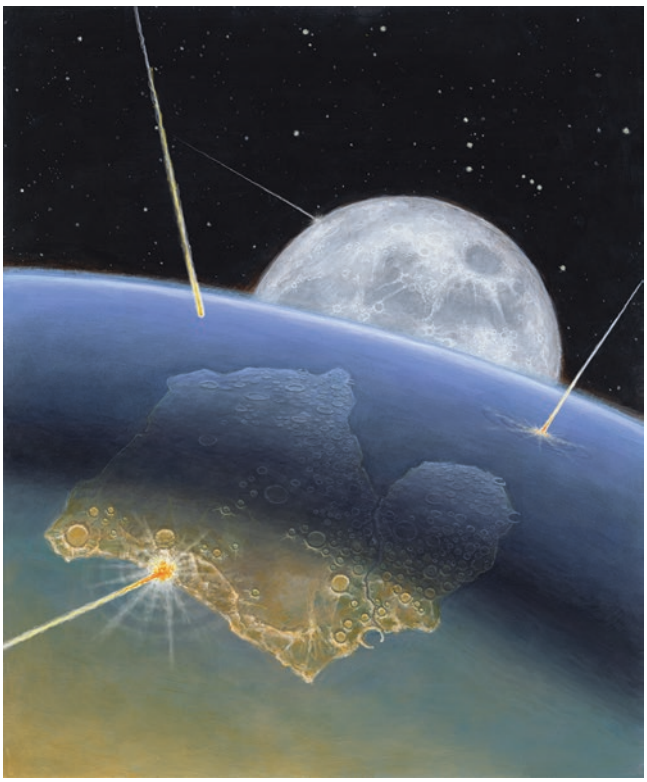


Fig. 6.1 Depiction of the full Moon rising over early Earth at twilight some four billion years ago. With a solid high-standing continental crust, the ancient supercontinent of Vaalbara was surrounded by the global ocean. The Moon was almost twice as close to Earth as it is today, filling much of the night sky. The surface of Vaalbara was as intensely cratered like the Moon. During the Eoarchean Eon, meteorites brought the building blocks of life to early Earth. They also created innumerable craters with hydrothermal systems. In a short time, these craters filled with water, forming hydrothermal crater lakes. Cosmic prebiotic ingredients accumulated in these hydrothermal crater basins, promoting the synthesis of life on Earth

With the gravity of the planetary mass doing its part, Earth became a layered planet. The mantle began to circulate. When hot mantle plumes reached the seafloor as undersea volcanoes, they chemically interacted with the oceans and hardened into crustal rocks. The first oceanic crust was mainly basaltic, resembling our modern seafloor in composition. Today, the high-standing continental crust covers nearly a third of Earth's surface. The protocontinental crust, being lighter than the surrounding basalt floor of the oceans, was buoyant, and this explains why the contemporary continents are not submerged.

The Archean crust evolved from a highly mafic bulk composition before ~3 Ga to a felsic bulk composition by 2.5 Ga. This upper crust transition from mafic to felsic marks the onset of plate tectonics at ~3 Ga [4, 8]. By the onset of the Proterozoic Eon that followed the Archean (~2.5 Ga), the crust had already assumed much of its present makeup, and modern plate tectonic cycling began. The Proterozoic Eon was an extremely active period in Earth's history tectonically. The plate tectonic framework explains the origin of the Proterozoic granite. The Proterozoic continental crust mainly grew along the subduction zones, when the oceanic plate subducted beneath the protocontinental plate, generating magma to rise to the surface, creating coastal mountains, and expanding the continental area. A drastic change of crustal composition indicates that plate tectonics began during the Mesoarchean through the Neoarchean era, with sizable later additions to the continental crust occurring throughout the Proterozoic.

Unlike the Proterozoic Eon, there is no plate tectonic activity in the Eoarchean. There is little agreement about the processes by which Archean cratons formed on Earth initially surfaced with a oceanic basaltic crust. Nevertheless, although essentially reworked by younger orogens, recycled by plate tectonics, eroded, or buried beneath younger rocks, fragments of the Eoarchean crust have survived in the form of cratons, the primal cores of the continents. Cratons—old and stable parts of the continental lithosphere—include both the crust and the mantle. In contrast to their Proterozoic successors, Archean cratons exhibit crustal granite-greenstone terrains and are usually durable and buoyant. Eoarchean (4.0–3.6 Ga) crustal rocks that cover less than 1% of Earth's surface in the most ancient crustal nuclei of all continents yet discovered have received unprecedented attention from a broad spectrum of geologists because they are the relics of Earth's first continental crusts.

Up to 90% of the preserved Eoarchean cratonic crust is composed of tonalite–trondhjemite–granodiorite (TTG) series with enclaves of greenstone packages (basalt, komatiite, and hydrothermal sedimentary strata). The hydrothermal sedimentary strata of the greenstone belts have yielded the oldest microfossils. Eoarchean TTGs occur in several gneissic complexes worldwide and are mostly generated by

partial melting of hydrated rocks at different depths of variably thickened Hadean mafic crusts [9]. TTG is the significant component of Earth's oldest remnant continental crust and thereby holds the key to the origin of the continental crust and the start of plate tectonics. These scarce rocks have been identified in only 10 or so cratons around the world.

Although this association of TTG with greenstone enclaves is unique to the Archean, the origins of terrestrial cratons are still uncertain. Continental rocks are produced by a bewildering array of processes. Still, a critical step in their production is the partial melting of oceanic and mantle rocks in the presence of water. Granite forms in abundance only when the hydrated basaltic crust is subducted. Recent works have suggested that Earth's first continental crust was not formed by subduction and plate tectonics [4, 8]. There was no plate tectonic regime during the early Archean time; the lithosphere then was a single, solid but deformable shell. In the Eoarchean Eon, then, Earth's first stable continents did not form by subduction or plate tectonics. The triggering mechanism in the composition of the continental crust four billion years ago appears to be linked to a novel kind of geological regime. In the absence of plate tectonics, we need an alternative geodynamic model to explain the origin of the Eoarchean continental crust. A plausible alternative mechanism to plate tectonics in the Eoarchean appears to be an impact cratering process.

6.2.2 Impact Origin of Archean Cratons

The origin of Earth's fundamental crustal dichotomy between low-density continental crust and high-density oceanic crust before the onset of plate tectonics remains obscure. A likely mechanism of crustal modification during the LHB was impact cratering, a major geodynamic force during the Archean times for the creation of tiny clusters of granitic protocontinents by mantle upwelling [10–12]. These protocontinents were small, exposed, protruding islands surrounded by vast oceans. Impact cratering is now widely recognized as a ubiquitous geological process affecting all planetary bodies with solid surfaces [13, 14]. The cratered surface of the Moon validates large-scale impacts in its crustal evolution. Studies of these craters indicate that the most massive among them arose during the Eoarchean. It is evident that if an LHB occurred on the Moon, then Earth must have also been battered by an intense impact flux [14]. So far, no unequivocal record of the LHB on early Earth has been found because of the recycling and resurfacing of the Archean crust by erosion and plate tectonics. The oldest known impact structure from the Archean is the Maniitsoq Crater (~3 Ga) of the greenstone belt in West Greenland [15]. Most known impact structures date from the Paleoproterozoic, such as the Suavjärvi Crater in Russia (~2.4 Ga), the Vredefort Dome (~2 Ga) of South Africa, and the Sudbury Basin (~1.8 Ga) of Canada [16]. Still, inferences drawn from the Moon and Mercury of contemporary impact cratering events on Earth are strong and convincing.

The dominant impactors bombarding Earth, the Moon, and Mercury during the LHB were asteroids, not comets [13]. Bolide impact events were both more extensive and more frequent during the LHB than they are today. The terrestrial effect of the LHB lasted through the Archean eon, generating pools of melt rocks. The bolide impact hypothesis for the origin of Archean cratons requires a large bolide to pierce the thin Archean lithosphere and cause massive melting in the sub-lithospheric mantle [17]. Early large impact events could form a felsic crust on early Earth to form protocontinents; impact melt volumes exceed transient crater diameters greater than 500 km [12]. A large bolide that penetrates the crust would also trigger lithospheric thinning and mantle upwellings, creating a plume that would affect Earth's surface with massive partial melting of the sub-lithospheric mantle. The region of the excavated and thinned felsic crust would shorten and thicken to form a protocontinent [17, 18]. Thus, the initial basaltic Hadean crust would have been reprocessed to produce felsic crust pockets by crustal fractionation.

Today, continents form when one tectonic plate subducts beneath another into Earth's mantle, causing magma to rise to the surface. However, there was no plate tectonic regime during the early Archean. Therefore, Earth's first stable continents could not have formed by subduction [4, 8]. During the Eoarchean and Paleoarchean eras, impact cratering was the only credible tectonic force, and therefore it must have caused mantle upwellings, creating small clusters of granitic protocontinents, i.e., small islands surrounded by vast oceans. The Pilbara Craton of Australia and the Kaapvaal Craton of South Africa document the tectonic effects of massive impacts on continental crust formation. These cratons show extensional faults active about 3.47 Ga during felsic volcanism and coeval with impact layers [19]. Declining impact frequencies over time led to an abrupt transition to plate tectonics during the Mesoarchean. In summary, the impact cratering model offers a coherent geodynamic framework to explore the origin of the Eoarchean crust before the onset of plate tectonics.

Because of the low density of the granitic crust, the continental felsic crust, perhaps created by impact plume, began to segregate from the mafic crust in an island-like setting. Numerous small felsic bodies of these protocontinents protruding above the sea level emerged in the early Archean and began to grow by accretion. Frequent meteorite impacts created terrestrial impact crater lakes, ideal crucibles for the synthesis of life. Eoarchean rocks of TTG–greenstone assemblages appear in the oldest cratons, including the Nuvvuagittuq Craton of Canada, the Isua/Akilia Craton of

Greenland, the Kaapvaal Craton in South Africa, the Pilbara Craton in Australia, and the Singhbhum Craton of India [21]. They are composed of volcanic and hydrothermal sedimentary rocks intruded by granitoid. The hydrothermal chert beds in the younger sequences of these basins contain the oldest fossils of hyperthermophilic life (Fig. 6.1). These five major Eoarchean greenstone belts were the nuclei of the emerging protocontinents. They are valuable archives for understanding the paleoecology of the earliest life. Of these, the Nuvvuagittuq Craton of Canada and the Isua/Akilia Craton of Greenland were in the northern hemisphere and were proximate geographically, if not sutured together.

Similarly, in the southern hemisphere, the Kaapvaal Craton of South Africa and the Pilbara Craton in Australia clustered into one supercontinent, Vaalbara, as far back as 3.6 Ga [22]. The greenstone rocks between the Kaapvaal Craton and the Pilbara Craton show a close stratigraphic correlation. The evidence comes from lithostratigraphy, geochronological structures, paleomagnetism, impact ejecta layers, and microfossils. In Vaalbara, the Pilbara Craton is juxtaposed to the southwestern margin of the present-day Kaapvaal Craton (Fig. 6.2). The greenstone gneisses—a product of the deformation of the original granitic rock, combined with layered intrusions—provide a rare and useful reference model for early Eoarchean history. They represent the relics of the felsic protocontinents. Most likely, the Singhbhum Craton of India was proximate to the Vaalbara supercontinent.

Sediments in early Archean greenstones are largely cherts, and volcanic deposits erupted in shallow water in hydrothermal vent settings [22]. The newly formed crust of the greenstone belts of young Earth presumably withstood thousands of impact craters analogous to the cratered surfaces of the Moon and Mercury. Over the subsequent eons, weathering and plate tectonics entirely erased and resurfaced the pockmarked surfaces of the Eoarchean crust. However, the protocontinents of the greenstone belts in the Nuvvuagittuq Craton of Canada, the Isua Craton of Greenland, the Kaapvaal Craton of South Africa, and the Pilbara Craton in Western Australia were the target of frequent impacts during the LHB, creating thousands of terrestrial hydrothermal crater lakes that may have become the incubators for prebiotic chemistry and the origin of life. Thus, the LHB impacts created not only the continental crust but also the innumerable hydrothermal crater lakes where the first biosynthesis began. The first ecosystems emerged on these crater lakes on the Eoarchean continental crust, providing clues to life's earliest days on the planet [20].

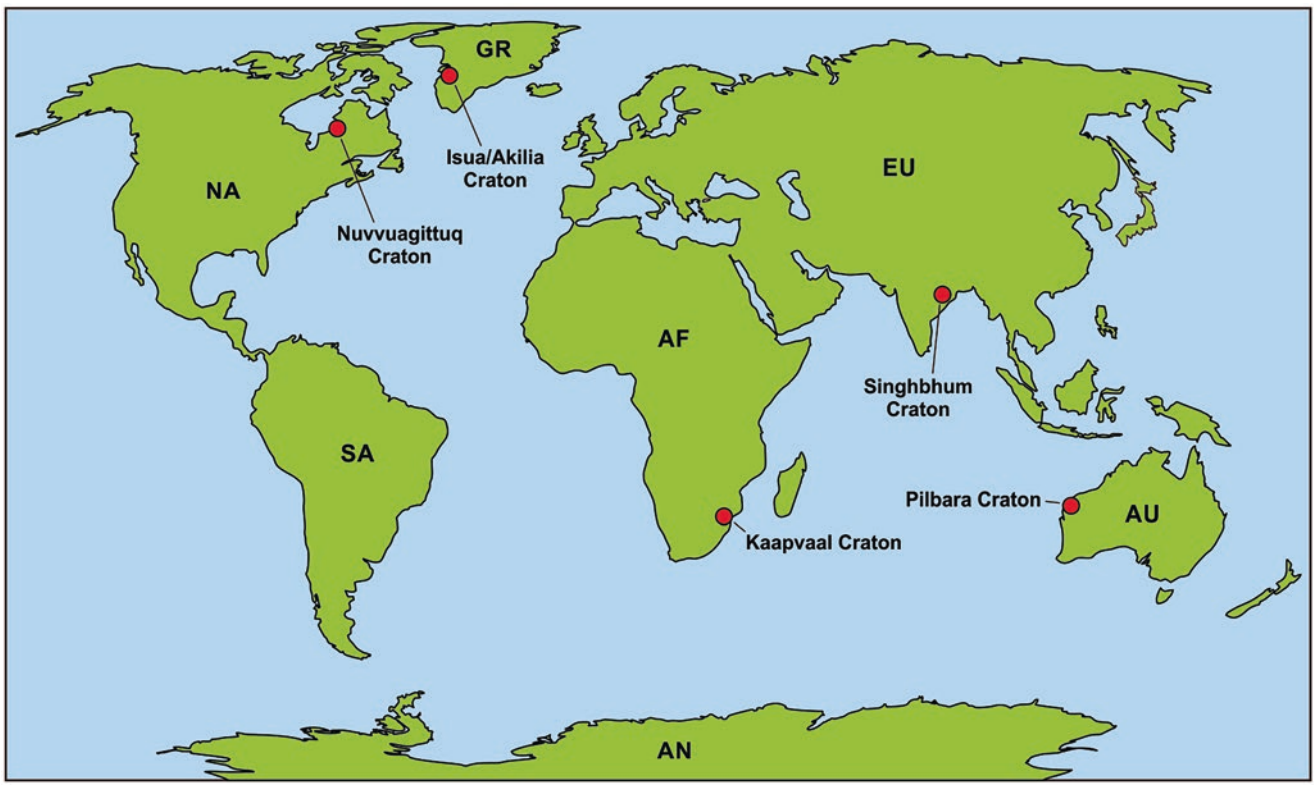


Fig. 6.2 The five locations in Archean greenstone belts where signs of the earliest life have been found: (1) Nuvvuagittuq Craton in Canada, (2) Isua/Akilia Craton in Greenland, (3) Kaapvaal Craton in South Africa, (4) Pilbara Craton in Australia, and (5) Singhbhum Craton in India. These greenstone belts are very likely relics of ancient impact scars on the protocontinents. Abbreviations: *AF* Africa, *AN* Antarctica, *AU* Australia, *EU* Europe, *GR* Greenland, *NA* North America, *SA* South America

Meteorite impacts could have generated both subaerial and submarine hydrothermal crater vents, but we prefer the subaerial setting for life synthesis. Terrestrial environments would have been more conducive to life than their submarine counterparts. The submarine hydrothermal crater vents in the vast global ocean environment would have been less likely to build up and sequester the concentration of biomolecules needed to start abiogenesis. We believe that contemporary submarine hydrothermal vents like black smokers and Lost City were colonized by life but did not model their original environments. However, such hypotheses deserve serious consideration alongside our preferred scenario favoring terrestrial environments for life's birthplace.

6.3 Habitat of Hyperthermophiles: A Window to Life's Beginnings

Several theories suggest a hot aqueous environment for the origin of life. The habitats of hyperthermophiles (superheat-loving microbes), the most primitive living organisms known, may shed new light on the oldest ecosystems on our planet—the cradles of life. In 1977, hyperthermophilic bacteria and archaea were discovered using chemical nutrients as a source of energy to thrive in deep, dark, anaerobic volcanic environments called submarine hydrothermal vents [23]. Today, hyperthermophilic bacteria and archaea have been found in numerous geothermally heated water-containing terrestrial, subterranean, and submarine high-temperature environments, such as hot springs and hydrothermal impact crater lakes, as well as the submarine hydrothermal vents along the mid-ocean ridge. They grow optimally above 80 °C and are viable up to 113 °C. Hyperthermophiles have certain heat-stable enzymes and rigid membranes adapted to high temperatures. In these hot anoxic environments, hyperthermophiles gain energy by redox reactions employing compounds like molecular hydrogen (H_2), carbon dioxide, sulfur, and ferrous and ferric iron. Hyperthermophiles were probably the most primitive living organisms on early Earth, about four billion years ago, and are still present today [24].

Because of their genetic antiquity, hyperthermophiles may provide a clue to the paleoecology and ancestry of life's beginnings. Carl Woese pioneered the application of a conserved molecule known as 16S ribosomal RNA (rRNA) to determine microbial relationships [25]. Molecular phylogeny suggests the division of living organisms into three domains of a 'universal tree'—Bacteria, Archaea, and Eukaryota. Moreover, all three domains descended from a single ancestral form at the base of the tree of life. This venerable ancestor of all life is called the last universal common ancestor (LUCA). New studies suggest that LUCA itself was a hyperthermophilic microbe [26]. There is no doubt that LUCA possessed DNA, RNA, proteins, a universal genetic code, ribosomes, and a proton-powered enzyme for making adenosine triphosphate (ATP) and lived in some hot, oxygen-free, and mineral-rich hydrothermal vent environment. The distribution of hyperthermophiles at the roots of Woese's universal tree suggests that hydrothermal systems, including submarine hydrothermal vents [27–29], tidal pools and hot springs [28], and hydrothermal impact crater lakes [16, 30–32], may have been the original site for life's beginnings. The paleoenvironment of the oldest sedimentary rocks containing evidence of early life (~4 Ga) supports this view [33–35]. Thus, hydrothermal systems may have provided favorable environments for the prebiotic synthesis of organic compounds necessary for life [36]. They represent the only known environment that could have created organic molecules with the same kind of energy-harnessing machinery as modern hyperthermophilic microorganisms. They could also have provided sanctuaries for the emergence of the first life during late, giant-impact events [20]. If the oldest fossils were indeed formed around four billion years ago, then the first life-forms would have had to 'like it hot,' as indicated by their paleoenvironment in the greenstone facies [20]. Most current theories of the origin of life suggest hydrothermal environments in oceans, hot springs, and impact crater lakes as the most promising sites for the origin of life.

6.4 Hydrothermal Systems and the Prebiotic Environment

Hydrothermal systems have persisted throughout geological history, and ancient hydrothermal deposits could provide clues to understanding Earth's earliest biosphere. Modern hydrothermal systems offer a dynamic, far-from-equilibrium reaction zone and a wealth of catalytic minerals allowing complex chemical reactions from simple organic compounds. They support a wide range of organisms and provide valuable comparisons for the paleontological interpretation of ancient hydrothermal systems. Continental and submarine hydrothermal systems are an integral part of Earth's thermal regime. Hydrothermal environments are dominated by the circulation of geothermally heated water and occur on both the ocean floor and the continents in areas with high heat fluxes. When water descending into the mantle meets the underlying magma chambers, its temperature rises well above the boiling point, but it remains in a liquid state due to the high pressure from the overlying water. Superheated water rises to the top, whereas colder, heavier water sinks around it. This creates convection currents that allow the more buoyant hot water to rise to the surface, following the cracks and weak areas through the magma source. Hot water rich in dissolved minerals from the underlying magma can penetrate cracks, fissures, and basin floors. We call this

downward and upward circulation the natural 'plumbing' system of hydrothermal features.

Geothermal vents are geochemically reactive habitats that harbor diverse chemosynthetic ecosystems and derive energy from chemicals and volcanism, not from sunlight. Microbial life thrives in the superheated water of these hydrothermal sites. Variations in the temperature and density of fluids drive the convection circulation in the crust, producing large-scale transfers of energy and material. The discovery of hydrothermal systems revealed vast and previously unknown domains of microbial habitats on Earth. In these extreme environments, the microbial and geothermal interactions are tightly interconnected, providing many essential constituents for the prebiotic synthesis of organic molecules and the evolution of fundamental metabolic processes of some modern prokaryotic chemoautotrophs. The biochemistry of these modern hyperthermophiles may well provide clues about the kinds of reactions that initiated the chemistry of life.

6.5 Geologic Setting of Hydrothermal Systems

Hydrothermal systems can develop anywhere on the crust where water coexists with a magmatic heat source. Such hydrothermal environments are likely to have encouraged chemical evolution and the subsequent origin of life. We believe that life started in some hydrothermal environment, whether at the bottom of the oceans or at the top of volcanoes. Among various hydrothermal settings, three locations stand out as the possible incubators of life: (1) submarine hydrothermal vents, (2) hot springs and geothermal fields, and (3) terrestrial hydrothermal crater lakes. There are many controversies about the most likely site for life's beginnings, and each prospective environment has its pros and cons. Here, we will evaluate these three possible hydrothermal sites one at a time.

6.5.1 Submarine Hydrothermal Vents

It was once believed that the ocean floors were dark, cold wastelands with little if any life. That image shattered in 1977 when oceanographers took the submersible vehicle, 'Alvin,' to the Pacific Ocean floor 3 km below sea level near the Galapagos Islands to study ridge volcanism where the tectonic plates diverge. Here, cold water descends through the crust into the underlying geothermally heated regions along the mid-ocean ridge axis. Hot chemical-rich water wells up from below the seafloor and pumps out through crustal cracks, creating chimney-like structures 8–15 m high. Named 'black smokers,' these chimneys spew steam water up to 400 °C, with high levels of sulfides and other minerals. Upon contact with the cold ocean, these mineral loads precipitate to form black smoker chimneys. Appearing through 'Alvin's' portholes was a previously unknown ecology. Beyond the black smoke issuing from the hydrothermal vents, an unimagined oasis of microbial and animal communities flourished in the dark, hot, oxygen-depleted rocky cracks and crannies of this previously obscure environment. An abundance of life apparently prospered in conditions without sunlight once considered too extreme to harbor much of any life at all.

To maintain life, these hydrothermal systems must be open, accepting energetic input continually or periodically in a nonequilibrium state. In fact, these vents are bursting with energy. Moreover, the microbial gardens at the hydrothermal vents have raised an exciting possibility: could these vents have provided the nutrients, energy, and protection necessary to incubate and eventually sustain early life? It is now clear, at least, that volcanic heat and exothermic reactions could drive the circulation of nutrient-rich fluids from which chemosynthetic microbes gain the metabolic energy to make a living along with dense populations of bizarre life-forms, including tubeworms, ghostly fish, and eyeless shrimp. The hyperthermophiles of this alien world support diverse marine communities of gastropods, bivalves, crustaceans, fish, clams, and limpets. Subsequent explorations have found hydrothermal vents surrounded by large populations of strange sea life along the mid-ocean ridges worldwide. These findings soon inspired proposals that submarine hydrothermal vents could have served as ideal environments with all the necessary ingredients for the first life to emerge [27].

The discovery of black smokers was followed in 2000 by yet another deep-sea find, a type of alkaline hydrothermal vent that occurs some distance away from the mid-ocean ridges, with cooler temperatures ranging from 45 to 90 °C. Dubbed the Lost City, this field of alkaline hydrothermal vents is unique among the seafloor hydrothermal systems, in that it is not driven by the cooling of deep magma or by seawater interacting with hot lava. Instead, at Lost City, hydrothermal activity is driven by chemical reactions between seawater and upper mantle rocks. These include layers of peridotites formed by a process called 'serpentinization,' which occurs in slow-spreading ridge environments where mantle rocks are commonly exposed near the seafloor. Serpentinization creates exothermic reactions that produce heat energy. It also yields alkaline vent fluids, sometimes called 'white smokers,' which rise from the seabed and build tall, castle-like enchanting structures. Named for these collections of spires made entirely of pure carbonate, Lost City was discovered on the seafloor by the Atlantis Massif in the mid-Atlantic Ocean. Both white and black smokers coexist in the same hydrothermal field, but they appear at distal and proximal vents, respectively, to the main upflow zone (Fig. 6.3).

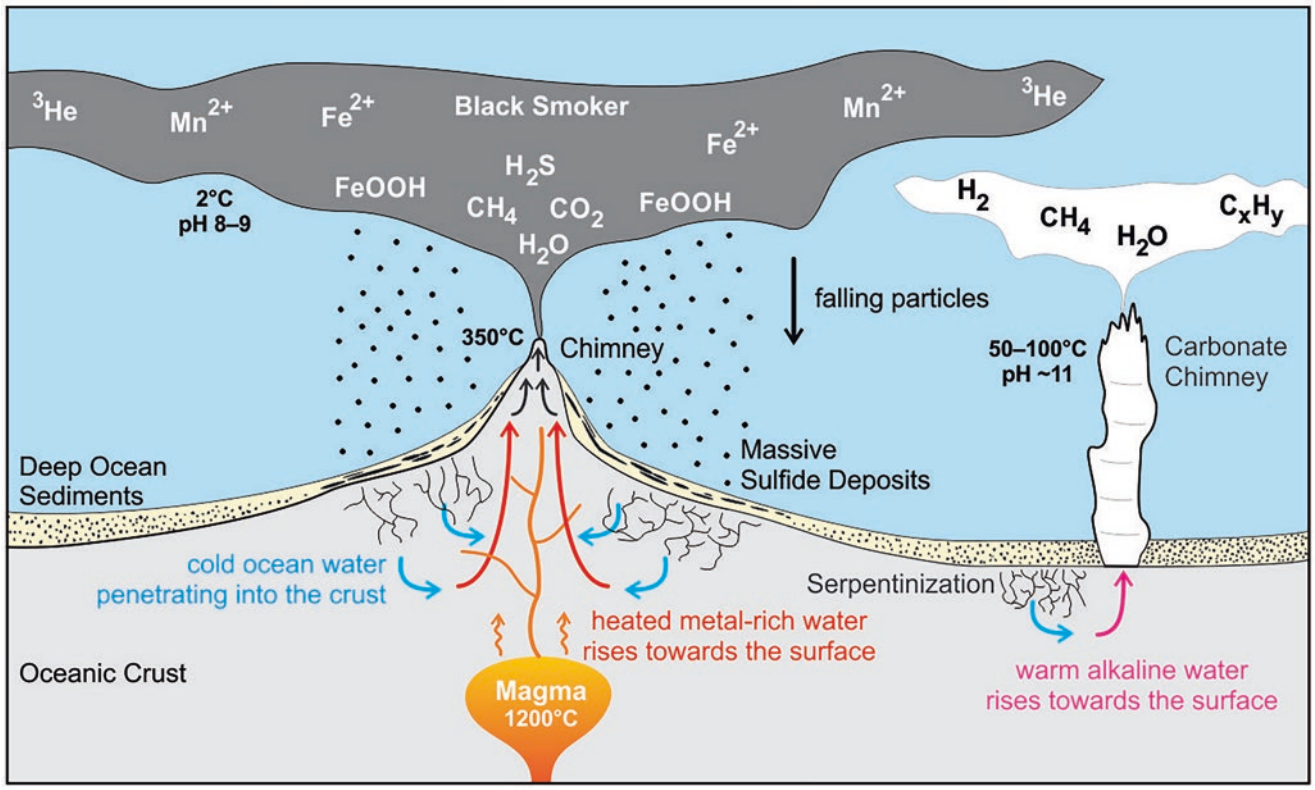


Fig. 6.3 A submarine hydrothermal vent environment showing discharges of the most abundant carbon compounds, toxic gases, metals, and energy sources. There are two kinds of submarine hydrothermal vents: left, the hot (~350 °C) black smoker, the chemistry of which is driven by the magma chamber that resides below the ocean floor spreading axis; right, the cooler (~50–90 °C) Lost City-type located far from the ridge axis, where serpentinization drives the chemistry

There is a considerable consensus that submarine hydrothermal vents were the most likely habitat for life's origin and early evolution [27–29]. They harbor the ingredients to fuel many of the chemical reactions necessary for the emergence of life. Of course, the ocean floor is too deep for sunlight that drives photosynthesis to penetrate. Instead, organisms here survive by chemosynthesis, deriving energy from chemical reactions. On the ocean floor, hydrothermal vents form in volcanically active areas—often on or near mid-ocean ridges.

Along with gases spewing from the hydrothermal vents, the cosmic building blocks would have entered a complex series of far-from-equilibrium chemical reactions in a scorching, dark, and highly reducing environment. For all their extreme temperatures, pressure, and other properties, deep-sea vents may have offered a relatively safe haven on early Earth's violent landscape for the emergence of life. In addition, the processes proposed for hydrothermal vents have also informed ongoing debates about whether, during prebiotic synthesis, metabolism or nucleic acids came first. The potential for prebiotic metabolic reactions in vents with substantial catalytic minerals supports the idea that basic metabolisms were the first to be established and then encoded later in genetic materials.

6.5.1.1 Black Smokers

Hydrothermal circulation is a typical process along the spreading center of the mid-ocean ridge. This subsurface circulation is manifested by the hydrothermal fields on the seafloor, which become the hubs of submarine oases of life. More than 500 such sites have now been located along the mid-oceanic ridges, spewing thick smoke-like plumes of black metal sulfides. At the mid-ocean ridges, the tectonic plates spread apart, magma wells up to the surface, and intense geothermal heating generates hot springs on the seafloor. At a hydrothermal vent, seawater seeps through cracks, dissolves metals and minerals, becomes superheated from the upwelling magma, and erupts like a geyser. These eruptions lead to the formation of black smokers [27, 29, 37–39]. During the cycling of seawater along the vent's axis, geothermal energy is transformed into chemical energy in the form of reduced inorganic compounds. When water rich in dissolved minerals from the crust, most notably sulfides, meets cold ocean water, minerals precipitate to form a black, chimney-like structure around each vent. The spewing water of these black smokers can reach temperatures of 400 °C. The vent supplies acidic fluids (pH 2–3) and dissolved reduced gases (hydrogen sulfide (H_2S), H_2, CH_4, CO_2, and ammonia

(NH_3)) that facilitate the chemosynthetic reactions of microbes that have evolved highly efficient ways to use geothermal energy. As they emerge from deep within the mid-ocean ridges, after the rising water, they are dissolved in cools and mixes with seawater; many chemical entities are not in chemical equilibrium. As a result, they are ready to enter into chemical reactions. Hyperthermophiles consume these compounds and use enzymes to speed up chemical reactions, releasing the energy that maintains their metabolism. In a vent environment, the predominant energy and carbon sources that microbes use originate de novo due to hydrothermal processes. Some microbes use CO_2 as their primary carbon source; others utilize CH_4. Ecologically considered, these hyperthermophilic prokaryotes, bacteria, and archaea are producers or autotrophs.

At an average depth of 2100 m, black smokers provide dynamic far-from-equilibrium systems and a wealth of catalytic minerals that could have been harnessed for the origin of life. However, although they provide essential conditions for primitive hyperthermophilic life, the most fundamental problem that such hypotheses confront in high-temperature and low-pH conditions is the lack of molecular stability during the emergence of life. In this condition, the bonds that hold together vital cell components, such as nucleic acids, proteins, and membranes, are prone to be torn apart as quickly as they are produced during the prebiotic stage, stopping biogenesis in its tracks [40]. However, once these hyperthermophiles appeared, their special high-temperature enzymes and tough membranes developed a viable level of thermostability [41]. This thermal paradox suggests that the crucible of life was cooler than the black smoker environment. Moreover, black smokers are extremely hot and acidic and short-lived, lasting only a few decades [42]. The origin of life is more fruitfully approached as a long step-by-step process in a stable location rather than as a series of low-probability events occurring at ephemeral sites.

6.5.1.2 Lost City

An alternative view is that life might have originated in alkaline, off-ridge submarine hydrothermal vents—these giant white carbonate chimneys are called Lost City for their spectacular spires of rock. The spires develop from the exothermic reaction of seawater with minerals resulting from serpentinization, such as olivine, and releasing white carbonates [28, 42–44]. Unlike the black smokers that directly reside on the seafloor spreading zone, Lost City is located several kilometers away from the mid-ocean ridge; it is highly alkaline (pH 9–11) and has a relatively cooler temperature (40–90 °C) than that of black smokers (Fig. 6.2). Unlike black smokers, the chimneys of Lost City, which may reach 8–60 m, are perforated with a labyrinth of interconnected microspores, like a mineralized sponge. Alkaline hydrothermal fluids percolate and potentially provide the source of a natural proton gradient [45]. Thermal currents through microporous labyrinths could have concentrated organic molecules (such as amino acids, fatty acids, and nucleotides), thus promoting prebiotic synthesis.

Lost City ecology is also a diverse array of hyperthermophilic bacteria, methane-cycling archaea, and other micro- and macroorganisms. The highly porous Lost City chimneys would also have provided favorable conditions for many simultaneous chemical reactions in separate microchambers, possibly leading to prebiotic synthesis and the first life on Earth [28]. However, the vent fluids of the Lost City system are extremely different from those of black smokers. Lost City vents release methane, ammonia, hydrogen, calcium, and traces of acetate, molybdenum, and tungsten into the surrounding water; they do not produce the major outputs of black smokers, large amounts of carbon dioxide, hydrogen sulfides, or metals. These vent systems are powered by serpentinization, the reaction of water with olivine, producing the hydrated mineral serpentinite and releasing warm alkaline fluids rich in hydrogen. The Lost City system may have been active for 120,000 years, significantly longer than black smokers. With substantial free energy in sustained far-from-equilibrium conditions, Lost City chimneys have been proposed as likely sites where chemical reactions could have initiated primitive metabolism involving the reduction of carbon dioxide by dissolved hydrogen [42].

Serpentinization is an exothermic reaction and releases heat. The process of serpentinization provides heat to drive the Lost City hydrothermal system in two ways. First, the residual heat of the mantle rocks underlying Lost City can be tapped through cooling with seawater. Second, the alteration reactions themselves are an important heat source:

$$\begin{aligned}\text{Olivine} + H_2O + CO_2 \rightarrow\ &\text{serpentine} + \text{magnetite} + \text{brucite} \\ &+ CH_4 + H_2 + \text{hydrocarbons}\end{aligned}$$

The formation of magnetite during the serpentinization process involves the oxidation of ferrous iron in olivine to form ferric iron in magnetite. Consequently, serpentinization produces reduced gases such as hydrogen, methane, and hydrogen sulfide. These dissolved gases provide essential energy sources for Lost City's microbial activity. Thus, the basement rocks and porous Lost City structures bathed in volatile-rich, highly alkaline fluids create vital niches for many kinds of life. The metabolic strategies of organisms living in such hot, highly alkaline (pH 10–11) water is currently being studied. The white chimney's porous structure yields an extensive surface area colonized by archaeal and bacterial mats. Hydrothermal microbes include sulfur-oxidizing, sulfur-reducing, methane-oxidizing bacteria and

methanogenic and anaerobic methane-oxidizing archaea. Chemosynthetic methane- and sulfur-metabolizing communities dominate the Lost City hydrothermal ecosystems.

The processes at work in Lost City support an attractive hypothesis regarding the prebiotic energy flow across the rock walls before the inception of cell membranes. The reactants produced by serpentinization may be concentrated in pores within the vent chimneys. Each hydrothermal vent contains numerous pores that could have acted as natural experimental chambers for monomer assembly [43]. With the prebiotic invention of cellular pumps, these vents could have resulted in two chemical imbalances. First, the hydrogen-rich alkaline water from the vents meeting the acidic ocean water could have generated a natural proton gradient or chemiosmotic coupling within the pores of rocks as a potential energy source. Second, an electron transfer could have occurred when the hydrogen- and methane-rich vent fluid met the carbon dioxide-rich ocean water, generating an electrical gradient that could have led to the spontaneous formation of acetyl phosphate and pyrophosphate. These two molecules act just like ATP, the chemical that powers living cells today. Combined with the natural proton gradient from the vent, the porous alkaline rocks were ideal for the incubation of prebiotic metabolism. Once life had harnessed the alkaline vent water's chemical energy, according to this hypothesis, it started making complex molecules such as RNA. Eventually, it created its own membrane, became an actual cell, and departed the porous rock into the open water [28, 45].

6.5.2 Arguments Against Submarine Hydrothermal Vents

The submarine hydrothermal vent version of the origin of life hinges on the hypothesis that strong thermal and chemical gradients present near undersea vents can initiate prebiotic synthesis. There is a disagreement in the popular view that life began in deep-sea hydrothermal systems, and, as such, support of the submarine vent hypothesis has been waning in recent years [46–48]. Several problems associated with the deep-sea hydrothermal vent hypothesis may be insurmountable. Black smoker systems on the ridge axis are not only transient (in the order of decades) but also have violent flow rates and extreme temperatures. These are not ideal for abiogenesis. Both submarine hydrothermal vent theories suffer from the 'dilution problem' with organic compounds in the open sea, making it difficult or even impossible to concentrate either ions or organic molecular components [20]. Rather than being concentrated, the cosmic ingredients would be dispersed and diluted in the vastness of the Eoarchean global ocean before they can assemble into the complex molecules of life. Concentration and the 'crowding' of organic molecules are essential aspects of prebiotic chemistry. It is challenging to imagine how a sufficient concentration of reactants could occur in the open ocean. One crucial requisite for the origin of life is that comparatively simple biomolecules must get opportunities to form more complex molecules by the segregation and concentration of chemical compounds. However, it seems unlikely that cosmic and terrestrial chemicals could have been concentrated, mixed, selected, or organized in this way in the vast ocean. Second, in one-plate, pre-tectonic Eoarchean Earth without spreading ridges, it is challenging to explain how submarine hydrothermal vents could have formed in the first place. Today's hydrothermal vents occur along or near the spreading ridges of oceanic plates (Fig. 6.2), provided that plate tectonics did not begin before 3 Ga [4, 8]. However, if life began around 4 Ga, then we will have to seek alternative hypotheses if we wish to retain hydrothermal vent systems as likely crucibles of life's origin. We will need to look for these on land, not under the ocean [20].

Third, A.Y. Mulkidjanian et al. have discovered that the chemistry of modern cells provides important clues to the original environment in which life evolved [49]. They suggest that organisms have retained their chemical traits throughout the eons. It turns out that all cells contain a lot of phosphates, potassium, and other metals—but hardly any sodium. In contrast, seawater is rich in sodium but deficient in potassium and phosphates. The composition of a living cell did not match that of the ocean water then nor does it do so now. For example, the molecular backbone of RNA/DNA is made of phosphate; many ancient proteins require zinc, and the cytoplasm of cells needs potassium ions to link amino acids to create proteins. Seawater is deficient in these critical elements of life. The inorganic chemistry of cytoplasm rather mirrors that of freshwater pond and lake environments. These authors conclude that the first life began on land, not at sea. They argue that land-based geothermally active pools are the likely sites where magnesium, potassium, zinc, and phosphate are available to match the ionic content of cells. Freshwater hydrothermal vents are conducive to the lower salt and ion concentrations that allow fatty acid membranes to form. They speculate that cellular life may have begun in anoxic freshwater hot springs rather than in seawater and that key chemical pathways to ribonucleotides could be facilitated by ultraviolet (UV) radiation in such a subaerial landscape.

Fourth, alternating dry and wet conditions are necessary for the significant polymerization reactions that must occur during the emergence of life. One of the biggest arguments against submarine hydrothermal vents is that so many macromolecules—DNA, RNA, proteins, and lipids—are all polymers; they form by the wet-and-dry cycles of condensation reactions [46, 47, 50]. Such condensation reactions could not occur in deep-sea hydrothermal vents. The con-

tinuous supply of water puts thermodynamic limits on the condensation reactions required for the polymerization of monomers. However, wet and dry cycling occurs every day on continental hydrothermal fields. This allows for the concentration of reactants as well as polymerization.

Black smokers release acidic water rich in carbon dioxide, heated to hundreds of degrees Celsius, and packed with sulfur, iron, copper, and other elements essential to life. Currently, Lost City is favored over black smoker environments as the likely cradle for prebiotic synthesis. The seawater expelled here is highly alkaline and lacks carbon dioxide but is rich in methane and offers more hospitable temperatures. However, the Lost City hypothesis faces several challenges, too. First, Lost City fluids do not contain an abundant supply of catalytic metals (iron, copper, manganese, zinc, nickel, etc.), which are needed to polymerize biomolecules. Second, the oldest ecosystems on Earth, as preserved in the Eoarchean greenstone belts in Canada, Greenland, Australia, and South Africa, do not show any evidence of Lost City environments such as dominant carbonate deposits. Third, the first cells probably developed in zinc-rich environments; the cytoplasm in a cell is rich in potassium, zinc, manganese, and phosphate ions, which are not widespread in Lost City environments [49]. Fourth, nucleic acids, such as RNAs, are unstable in highly alkaline pH, raising the possibility that RNA must have arisen in a neutral or mildly acidic environment. This observation clearly weakens the hypothesis that RNA evolved in the vicinity of alkaline (pH 9–11) hydrothermal vents [51].

Our view supports shifting to a land-based origin-of-life hypothesis. The oldest fossil record from the Pilbara Craton of Australia also reinforces this approach. The fossil-bearing rock horizons of the Dresser Formation (~3.5 Ga) were formed in terrestrial hot springs, not in the ocean [52]. These concerns and criticisms lead to the consideration of alternative scenarios that life began in freshwater fields and were subjected to cycles of dehydration and hydration that promote polymerization. Many scientists now believe that the prebiotic synthesis could not have occurred in ocean settings and are looking for a land-based birthplace of life.

6.5.3 Terrestrial Hydrothermal Systems

Geological and chemical evidence increasingly supports the hypothesis that life originated on land rather than in a marine environment [49]. Hydrothermal basins are surprisingly common and widely distributed across the terrestrial surfaces of planet Earth. They have multiple geological origins, particularly in association with convergent plate boundaries and transform faults. However, since plate tectonics had not begun at the time of the Eoarchean era, we can discount these tectonic locations as plausible geological settings for the first life [4, 8]. The most relevant and numerous recent and active terrestrial hydrothermal environments are usually associated with volcanic features, which allow magma to penetrate upward to a shallow depth where it can directly interact with groundwater. While a magma chamber's depth may be as shallow as 2–5 km below ground level, the thermal gradients may be as high as 150–200 °C/km^{-1}. Here, the groundwater may be superheated and forced back to the surface at a high temperature and under considerable pressure. Most of these hydrothermal systems are populated by diverse hyperthermophilic microbial life. Terrestrial hydrothermal environments are widely considered suitable analogs to conditions that may have given rise to early life on our planet [24]. Terrestrial hydrothermal candidates for life's origin include (1) hot springs and tidal pools, like the central idea of Darwin's 'warm little pond,' and (2) impact crater lakes. Here, we evaluate both terrestrial environments as possible locations for life synthesis.

6.5.3.1 Darwin's Warm Little Pond Revisited

In February 1871, Charles Darwin wrote to his botanist friend Joseph Hooker speculating that life could have evolved in some 'warm little pond' if it were full of ammonia, phosphoric salts, light, heat, and electricity. In this small body of water, simple prebiotic compounds could undergo further reactions, thus producing more complex molecules. Descending but diverging from Darwin's early musing on the 'warm little pond' as a cradle for life synthesis, the marine origin theory for a life dominated the field for many decades, partly in response to Oparin's landmark work (see the first chapter). Oparin suggested that due to the combination of reducing gases and some form of energy such as lightning or UV rays, the primal ocean became a 'primordial soup' in which life could take shape. Since then, an oceanic origin of life has been the favorite explanation. The discovery of submarine hydrothermal vent ecologies in the 1970s reinforced the ocean as the likely incubator.

Recent studies have suggested that life on Earth may well have started on land in a 'warm little pond,' just as Darwin speculated about 150 years ago. Current views of the earliest fossil record suggest that the first life-forms inhabited terrestrial hydrothermal environments [33, 34, 52]. This terrestrial setting favors an origin of life in freshwater hydrothermal systems over the popular hypothesis of marine origin. Low-temperature hydrothermal sites exist where surface water penetrates sufficiently to be heated due to a normal crustal thermal gradient. Such hot springs frequently have a near-neutral pH and a temperature range of 40–60 °C. Despite their widespread occurrence across the terrestrial world, organisms that thrive there are thermophilic but not extremely so. The hyperthermophilic organisms relevant to the origin of life grow optimally at a higher temperature ranging from 80 to 110 °C. In submarine hydrothermal vents, the home of

living hyperthermophiles, the temperature can reach up to 340 °C.

Comet Pond

Another variant of the warm little pond scenario is the comet pond model, which envisions the landing of a relatively pristine portion of a comet containing building blocks of life on a planetary surface during the LHB, by soft impact, thus creating a shallow pond to serve as an incubator. In this pond of highly concentrated organic-rich matter, prebiotic synthesis would have started up when ultraviolet radiation polymerized the cometary monomers, eventually leading to the origin of life on Earth [53]. Although the comet pond theory is entirely novel and plausible, the authors acknowledge that because of the rapid destruction and dispersal from entry through the atmosphere or a hypervelocity impact during landing, the delivery of pristine cometary material would be an improbable event. Moreover, it has been argued that the LHB was exclusively composed of asteroids rather than comets [54].

Volcanic Hot Springs

David Deamer of the University of California and his associates have championed an updated version of Darwin's warm little pond, focusing on fluctuating hydrothermal pools such as terrestrial hot springs [30, 46, 47]. The setting for this proposed model of a hydrothermal site is a volcanic island emerging from a global ocean four billion years ago. Several modern analogs of hot springs in volcanic islands at the time of life's origin have been proposed, including Bumpass Hell hydrothermal fields in Lassen Volcanic National Park in California, Mount Mutnovsky in the Kamchatka peninsula of Russia, and sites in Iceland and Hawaii. In many thermal hot springs, precipitation from rain and snow percolates through porous rock and, once deep underground, is turned into steam by magma. Although most subterranean steam condenses below ground, some reach the surface through conduits to form fumaroles, hydrothermal fields, and hot springs (Fig. 6.4). Primordial hot springs would have been able to concentrate organic molecules from meteoritic sources and interplanetary dust particles to initiate prebiotic synthesis. Because these hydrothermal ponds tap freshwater from rain and snow, their ionic concentrations are much lower and so more favorable. The chemical composition of volcanic pools also resembles the composition of cytoplasm more closely than that of the submarine vent environment. In addition, small ponds undergoing evaporation and replenishment by variable hot springs and precipitation go through fluctuating cycles of hydration and dehydration.

Subaerial hot springs offer several advantages over submarine hydrothermal vents as sites for abiogenesis. Three key features of these environments are conducive to prebiotic synthesis. First, hydrothermal fields provide sources of heat and chemical energy for prebiotic synthesis. Second, hydration–dehydration cycles drive condensation reactions, producing long-chain organic polymers by a nonenzymatic reaction. These pools dry up periodically. Deamer imagines volcanic islands with freshwater pools sprayed by hot

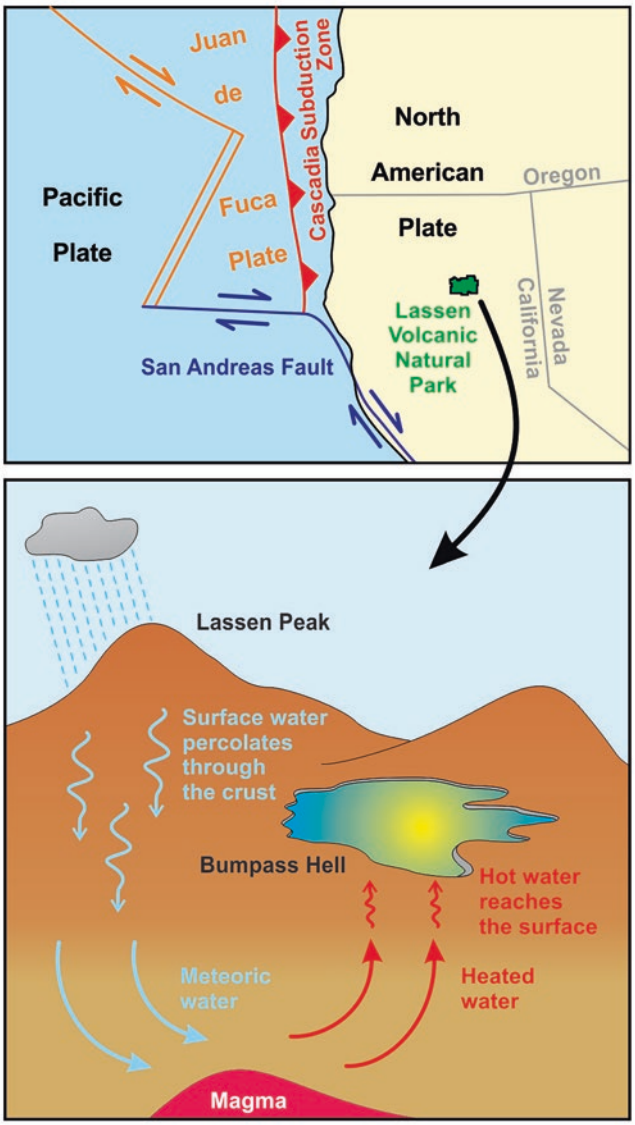


Fig. 6.4 The modern analog of Darwin's 'warm little pond' such as the terrestrial hot spring exemplified by Bumpass Hell at the Lassen Volcanic National Park, California. Top, subduction of the Pacific plate under the North American plate along the Cascadia Trench created the Cascade Volcanic Range from British Columbia through Washington and Oregon to Northern California, a distance over 1100 km; Lessen Peak erupted from 1914 to 1921. Bottom, the underlying magma source at Lessen National Park is the manifestation of this Cascadia subduction; water from rain and snow that precipitates on the highlands of the Lassen Volcanic National Park feed the Lassen hydrothermal system. Most of the hydrothermal features contain mixtures of the condensed steam and near-surface groundwater. However, in the Eoarchean, when life might have emerged, there was no plate tectonic activity to create this kind of hydrothermal system at the trench boundary

springs or geysers in between drying out. These pools would have been much more apt locations than deep-sea vents to form the first fatty membranes or amphiphiles. In the hydration phase of a cycle, vesicles are produced when the water interacts with the dry multilamellar matrix, whereas during the dehydration phase, solutes and amphiphiles on mineral surfaces form highly concentrated films of multilamellar structures that capture concentrated monomers between layers. The amphiphiles and minerals provide environments for concentrating solutes and for driving condensation reactions that link monomers to polymers. Third, self-assembly of membranous compartments would encapsulate the polymers into populations of protocells, leading to stepwise increments of chemical evolution toward the emergence of first cells. Moreover, the 'warm little pond' theory has supporting evidence from several experiments, rendering competing for the submarine vent hypothesis more problematic [46, 47, 49].

However, there is an issue with this volcanic hydrothermal field scenario from a geological perspective. The modern analogs provided by these authors were created by plate tectonic activity—the Bumpass Hell formed by the subduction of the Pacific (Juan de Fuca) plate under the North American plate along the Cascadia Trench (Fig. 6.4) and the Kamchatka Peninsula formed along the subduction of the Kuril–Kamchatka Trench at a triple plate collision between the North American, the Pacific, and the Eurasian plates. Moreover, Iceland sits right on the mid-Atlantic Ridge under the North Atlantic Ocean, where the Eurasian and North American plates slide apart. The volcanoes caused by subduction or spreading ridges are powered by the same magmatic heat that creates many modern hot springs. Since plate tectonics did not start until after the Eoarchean period, when life began [4, 8], these contemporary examples are not relevant for abiogenesis.

Similarly, Hawaii's volcanic islands result from hotspots beneath the North Pacific Ocean. In the absence of plate motion, volcanic islands would be solitary instead of a chain. Isolated volcanic islands would have eventually eroded, submerging their 'warm little ponds,' releasing their diverse biomolecules and dispersing them into the ocean, thus halting further abiogenesis.

Other researchers have also elaborated the volcanic island model, suggesting as a modern analog for prebiotic synthesis a site such as Lake Waiau, about 3970 m above sea level near the summit of Mauna Kea volcano on the island of Hawaii [55]. However, this lake cannot retain water and sometimes completely dries up in the summer and freezes in the winter. The lake's seasonal temperature is too low (~0–13 °C) to support hyperthermophilic life, presumably the first lifeforms. From a geological perspective, then, the volcanic scenario is problematic. Most of the islands during the Eoarchean were continental, not volcanic.

Settings such as the Yellowstone formed by the interaction between a hotspot and the south-westerly drifting North American plate is another example of terrestrial hot springs. However, the duration of hydrothermal activity at the Yellowstone National Park ranges from a few hundred to a few thousand years, apparently too brief a time for abiogenesis to take place [56]. Moreover, without plate tectonics, the plume would be short-lived.

Geothermal features are also observed in areas of active volcanism, other than plate tectonics, where subsurface magma heats groundwater, creating hot springs. Hot springs can also form along the active fault zones, where the groundwater is heated by circulation through faults to the rock deep in the crust. In terrestrial settings, perhaps this kind of hot springs come close to Darwin's warm pond scenario as championed by David Deamer.

Asteroid Ponds

Pearce et al. [50] suggested a hybrid model of Darwin's warm little pond when an impacting carbonaceous chondrite broke into smaller fragments that crashed into terrestrial pools and released soluble nucleobases, such as adenine. The wet-and-dry cycles of the pond would promote condensation and rapid polymerization of nucleobases into RNA. These authors provided a quantitative estimate that RNA-like polymers could be synthesized in just a few wet-and-dry cycles rather than many million years. However, meteorites splashing into warm little ponds would likely vaporize the water instantly, halting any biogenesis at hand. However, in fact, this scenario is a modified version of the impact crater lake hypothesis with a pond-sized crucible allowing wet-and-dry cycles. Other than wet-and-dry cycles, mineral surfaces on the crater floor are equally effective in polymerization [26, 29, 32, 48]. Statistically speaking, during the Late Heavy Bombardment of the Eoarchean crust, impact crater lakes were more common than were warm little ponds containing meteorite fragments.

6.5.3.2 Impact-Generated Hydrothermal Crater Lakes

Impact Craters

A unique combination of circumstances on early Earth enabled the entry of pristine building blocks of life aboard incoming asteroids and their landing onto an Eoarchean crust to create myriads of hydrothermal crater lakes. The most significant advantage of the cratering hypothesis for the origin of life is that it invokes a scenario of high probability. Because of the Late Heavy Bombardment, impact cratering was the most common landform on the Archean crust. Impact craters are also of high interest in planetary exploration because they are considered possible sites for life evidence.

Compared to the occurrence of such events on present-day Earth, during the Hadean–Archean transition, Earth's

surface was subjected to relatively heavy bombardment by asteroidal impacts [10, 13, 31, 32, 54, 57]. If the pockmarked surfaces of the Moon and Mercury are any indication, then early Earth was pelted by asteroidal swarms for about 100 million years (see Fig. 5.3). However, because our planet's surface is constantly eroded by the hydrosphere and renewed by plate tectonics, the physical evidence of that early bombardment has been wholly erased. However, since it roughly coincides with the time, we believe that the first life appeared on our planet, and this early period of heightened asteroidal activity would have had significant implications for life on Earth. A barrage of asteroids punched through Earth's crust, creating thousands of impact craters, delivering enormous thermal pulses to local environments riddled with volcanically driven geothermal vents. When the impactor's kinetic energy is transferred onto a water-bearing crustal surface, the resulting crater exhibits long-term hydrothermal activity. When filled with water, these crater basins developed hydrothermal systems that could have helped sustain the first life by providing sources of heat, energy, water, and nutrients. While the impact-generated silicate debris in the crater basins also produced clay minerals, the hydrothermal fluids precipitated pyrites, components likely to have played critical catalytic roles in the organic synthesis of prebiotic and early biotic compounds and structures [48].

The interaction of terrestrial melt rock with groundwater generated a hydrothermal system. Rainwater eventually filled the crater to form a lake, and the heated floor of the crater sustained the hydrothermal system for an extended period. The crater rim rose considerably above the lake surface and formed an ideally isolated basin for the concentration of biomolecules. Impacts on a water-rich planet, such as Earth or even possibly Mars at that time, could have generated hydrothermal activity in which underwater areas boil with heat. The central peak of a complex crater (> 5 km) might have been fractured by such an impact, providing a high geothermal gradient, and spewing forth chemicals, as occurs along the axes of submarine hydrothermal vents. The central peak may have been an essential source of heat in the creation of a hydrothermal system. We think of cooling impact craters as natural laboratories where convection currents circulate hot hydrothermal fluids. Impact-induced hydrothermal systems are highly diverse, with a wide range of pH and temperature gradients. There is a temperature gradient inside the crater basin, where the surface water is cold. Still, the region adjacent to the central peak on the floor is hot, creating a steady convection current that helps mix the cosmic ingredients. Depending on the impactor's size, they would cool down at different rates, thereby generating hydrothermal systems of various sizes and temperatures, creating a variety of incubators with complex compounds undergoing different chemical reactions. The larger the crater, the longer its hydrothermal activity can persist. While a small crater with a diameter of 5 km can sustain a hydrothermal system for several thousand years, a crater of 200 km diameter can do so for hundreds of thousands of years. During the lifetime of a crater, the chemical reactions that may have facilitated prebiotic synthesis can vary. During an early hot stage, simple cosmic molecules could be concentrated; in later stages, lower temperatures would be conducive to more complex molecules in different regions of the impact structure.

Ironically, with the extinction of the dinosaurs, the public associates meteoric impacts with death. However, they could have also brought the building blocks of life and water to Earth. However, subsequent meteor collisions may have slowed down abiogenesis or even sterilized pockets of the first life. The rhythm of creation and destruction is not only manifest in the origin of life and the birth and death of all living creatures, but is also the very essence of the evolution and extinction of life. Like Shiva's cosmic dance, meteorites can be both creative and destructive. Thus, the underwater havens of the impact crater lakes, especially the impact-shocked, porous rocks on the basin floor, could have provided a crucial sanctuary for life's origin, protecting the first microorganisms not just from the Sun's harsh ultraviolet rays when the planet still lacked the ozone shield but also from the continued violence of the Heavy Bombardment. That is why we believe that these crater lakes were the perfect locations to concentrate and cook these cosmic chemicals to create the first simple microbes [16, 20, 32, 57].

Hydrothermal Systems in Crater Lakes

Abundant on the Eoarchean crust, hydrothermal systems in numerous crater lakes are the likely cradles for the origin of life. Derived from meteorite impacts, the building blocks of life began to concentrate here. In addition, a continuous infall of micro-meteorites containing iron, manganese, and silicates formed a blanket around the lifeless surface of newly formed crater lakes. It started to interact with biomolecules in the basins. Hydrothermal vents inside the crater basins of these sequestered hydrothermal crater lakes provided a continuous stream of energy to mix chemical compounds with cosmic ingredients.

Hydrothermal chemistries were powered by solar, tidal, and chemical energies, including ATP [62]. Convection currents formed a sticky, brownish, primordial prebiotic soup in which concentrated cosmic and terrestrial chemicals rich in organic compounds could grow into larger and more complex molecules [20, 57]. Hydrothermal impact crater lakes provided for the selection, concentration, and organization of specific organic molecules into successively more information-rich biopolymers and finally into the first cells [59]. Moreover, if interconnected by underground fractures, then the neighboring impact craters of different sizes, temperature regimes, and concentrations could create a feedback loop among networks for abiogenesis.

The crater lakes possess all the previously proposed favorable characteristics of deep-sea hydrothermal vents for the synthesis of life, such as prolonged circulation of heated water for mixing and concentration of prebiotic soups; various chemicals, energy, and ATP for chemosynthesis; abundant catalytic surfaces of mineral substrates with nanopores and pockets for polymerization; and nutrients for primitive life. Moreover, each crater basin provides at a single location additional geochemical and environmental advantages over the submarine vents. These favorable constraints for life synthesis include [20]:

1. Impact events generate several habitats that are conducive to microbial colonization. The continental excavation by impacts, forming crater lakes, exposed a variety of rock types, including sialic, mafic, and ultramafic, which provided a dynamic fluid mixed with variable temperatures, magmas, and pH conditions (Fig. 6.5). This combination allowed an active fluid that favored the prebiotic synthesis in reducing and reactive habitats. However, at a purely volcanic and tectonic venue, such mixtures were likely to have been less available.
2. Whereas in the global ocean, most of the organic compounds from cosmic and volcanic sources would have been diluted and dispersed, inhibiting prebiotic synthesis, high crater rim sequestering, and compartmentalizing biomolecules would have been ideal for concentrating the complex reactions of cosmic molecules.
3. In contemporary niches, the chemical content of vent water originating from a complex set of reactions between lake water and hot, newly minted impact melt rock nourishes a chain of living microbial communities. It was likely to have done the same during biogenesis. Impact-induced fracturing on the floor of the crater basin increased the surface area for the concentration of polynucleotides and polypeptides [31]. As we have noted, hot water circulation from the central peak to the cold surface water creates a convection current that promotes the mixing and concentration of monomers and the widespread formation of clays and pyrites on the basin floor encourages the polymerization of monomers.
4. Terrestrial crater basins contained freshwater with low total salts and K^+/Na^+ ratios that match the cytoplasm of all cells in all three domains of life [49].
5. Hydrothermal reactions were the primary source of the metal-rich sediments and nodules that carpet the crater floor. The different metal sulfides, clays, pyrites, and zeolites that were available in crater basins would have functioned as catalysts for complex sequences of condensation reactions and polymerization of amino acids and nucleotides [31, 48, 57]. The 'iron–sulfur world' inside the hydrothermal vents promoted a variety of chemical reactions and acted as an early form of metabolism before nucleic acids appeared [29]. Mineral surface bonding and pore spaces are seen as functional precursors of the cellular enclosure and bonding enzyme surfaces.
6. Wet-and-dry cycles favored the concentration of biomolecules in the crater basin and condensation for polymerization at the exposed crater's edge [30, 46, 47].
7. High temperate ranges (~60–90 °C) and pH values near neutrality (pH 5–8) are ideal for stabilizing the membrane vesicles of emerging hyperthermophilic microbes [30, 46, 47].
8. Additionally, solar and tidal sources of energy were available for the evaporation and concentration of polymers [49].
9. Impact-shocked rocks and impact glasses offered endolithic habitats that would offer shelter against harmful UV radiation when Earth still lacked an ozone layer [32].

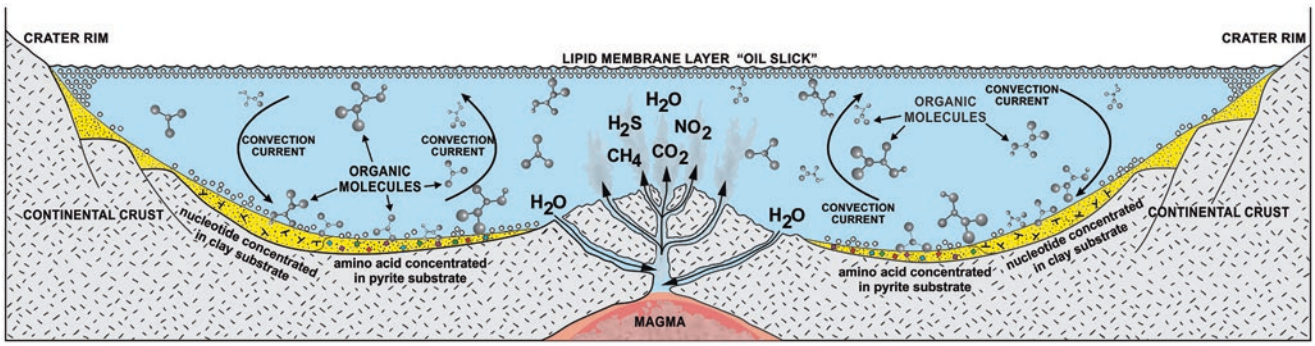


Fig. 6.5 Cradle of life. Hydrothermal crater lakes in the early Archean offered a protective haven for prebiotic synthesis. The boiling water was rich in the building blocks of life. On the surface crater basin, lipid vesicles and hydrocarbons were buoyant like tars. The mineral substrates on the floor of the basin acted as catalytic surfaces for the concentration and polymerization of monomers. Convection currents thoroughly mixed the bubbling biotic soup. Some lipid vesicles by convection current went down to the crater floor and stuck to the mineral substrate, encapsulating biopolymers such as RNAs and peptides. Hydrothermal vents provided heat, gases, and chemical energy, including ATP molecules

10. Because of the huge amount of energy from impacts, impacted melt rocks could keep the vent environments hot and reactive for long durations likely necessary for biosynthesis [32].
11. All hydrothermal crater lakes are abundant in H2S, a potential source of nourishment for emerging thermophiles. Even when hydrothermal activity ceased, hydrothermally altered and precipitated rocks would have provided enduring sources of nutrients and habitats [32].
12. When interconnected by subterranean fractures, impact crater lakes of variable sizes, each with a distinct set of features (such as pH, chemistry, temperature, and ionic concentration), could have mixed components and shared information during prebiotic synthesis (Fig. 6.6).

These 12 geochemical and environmental characteristics favor hydrothermal crater lakes as the likely site for life's origin rather than the submarine hydrothermal vent model.

Hydrothermal crater lakes encompass a multiplicity of physical, chemical, and mineralogical gradients shaping the structure of the microbial communities inhabiting these systems. The ecosystem of a hydrothermal crater vent includes a gradient of temperature, nutrient abundance, chemical environment, and pH, even within a single thermal regime. Because a wide range of environmental conditions were required to produce a living cell from organic precursors, the environmental complexity of the crater basin was a necessary requirement for the origin of biological complexity. The anoxic crater lakes of pre-plate tectonic Earth are unique in

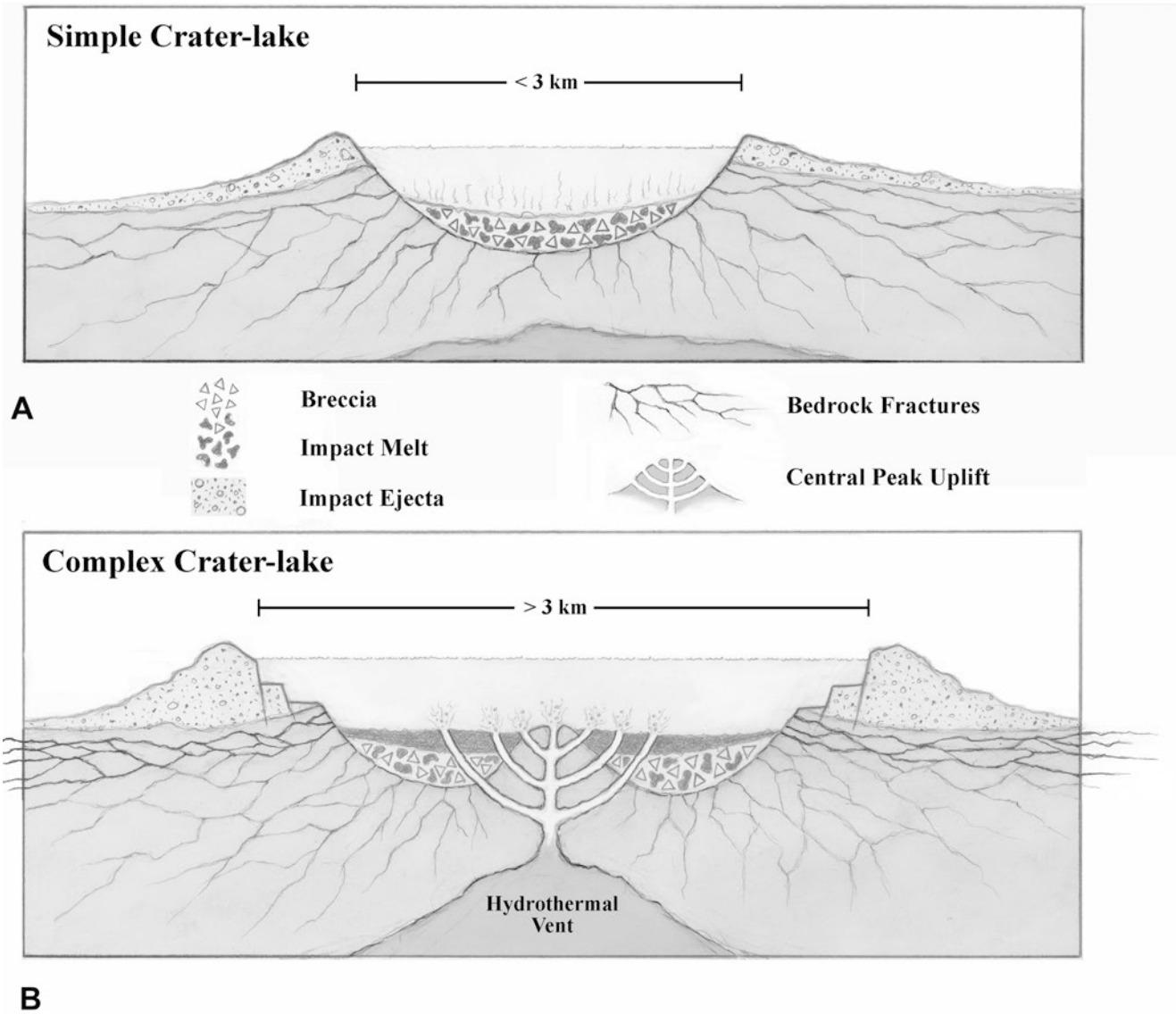


Fig. 6.6 Impact craters come in two groups based on their morphology: simple and complex. Complex craters (>3 km diameter) with central peaks generating a dynamic hydrothermal system would create ideal environments for the first crucibles of life

geological history and highly different from the submarine black smokers or the Lost City. The cold water of the crater lakes is heated by hot magma from the central peak and reemerges from the vents, reaching a relatively moderate temperature (40–90 °C). The fluctuating temperature gradient created a convection current within the lake water, mixing the assorted cosmic and terrestrial chemicals with reduced gasses such as H_2S, H_2, CH_4, CO_2, and NH_3. These mixed hydrothermal fluids formed a complex solution of thick, prebiotic soup. Alternating wet-and-dry cycles of the lake surface facilitated the concentration of organic molecules. They drove condensation reactions at the elevated crater's edge for the polymerization of monomers such as nucleic acids and amino acids. Mineral surfaces of the crater floor, such as clay and pyrite, could catalyze the polymerization of nucleic acid polypeptides and concentrate phosphate and polymers in pores and cavities for prebiotic reactions (Fig. 6.5).

6.6 Morphology and Distribution of the Impact Crater Lakes

Impact craters are formed when meteorites smash into a planet or a moon with a hard crustal surface. Meteors have heavily bombarded all the bodies in our solar system throughout their early period. The surfaces of Mercury, Mars, and the Moon have preserved cratering records of the Late Heavy Bombardment for billions of years because they lack plate tectonic activity. Crater formation is mediated by an energy transfer mechanism that transfers the kinetic energy of the impactor into heat, subsequently fracturing, displacing, and excavating the target rocks. Impactor velocity ranges from 10 to 70 km/s, with an average of 20 km/s on Earth. More than 150 impact craters have been identified on our planet in a wide diversity of biomes and with a variety of target rocks [63]. More than 70 of them show evidence of impact-generated hydrothermal activity, from the ~1.8-km-diameter Lonar Crater structure in India to the 250-km-diameter Sudbury impact structure in Canada.

Akin to hot springs and geysers, hydrothermal vents formed in impact-induced crater lakes. Impacts that excavated huge craters also shattered their central peaks to create volcanically driven geothermal vents in concert with groundwater. As rains filled up the crater basins, underwater hydrothermal vents developed and crater lakes formed. The reduced gases from hydrothermal vents mixed with cosmic biomolecules and thereby created ideal prebiotic mediums for biogenesis.

Impact craters are divided into two groups on the basis of their morphology: simple and complex [63]. Simple craters are relatively small, no more than 3 km across their uplifted rim, surrounding a bowl-shaped depression partially filled by breccia. Their maximum depth is located at the center (Fig. 6.6a). Complex craters and basins are generally 4 km or more in diameter; they have a central uplift peak surrounded by an annular trough and a slumped outer rim with relatively low depth/diameter ratios. Any meteorite impact capable of forming a complex crater lake can potentially generate a hydrothermal system (Fig. 6.6b). Simple and complex crater lakes sequestered by crater rims are both ideal sites for prebiotic synthesis. However, a small complex hydrothermal crater lake would enhance the concentration of biomolecules better than larger ones and would be even more ideal for the synthesis of life.

Terrestrial hydrothermal impact crater lakes have been proposed as the most likely sites for biogenesis on Earth [16, 20, 31, 32, 57] and, by analogy, on Mars [64]. These crater lakes could have also provided a refuge for the earliest hyperthermophilic life during late, giant-impact events. Craters were presumably plentiful on the Eoarchean crust; this period also overlaps with the evidence of the earliest life on Earth. Moreover, modern subaerial hydrothermal lakes are widely colonized by hyperthermophilic, thermophilic, and mesophilic bacteria and archaea.

Due to the shift of kinetic energy from the impactor to the target, every impact on water-bearing crustal surfaces generates some long-term hydrothermal activity in the resulting craters. Impact events trigger shock pressures and temperatures that can melt substantial volumes of the target material. The three primary potential sources of heat for creating impact-generated hydrothermal systems are as follows [32]:

1. Impact melt rocks and impact melt-bearing breccias
2. High geothermal gradients occurring in central uplifts
3. Energy deposited in central uplifts due to the passage of the shock wave

Complex craters with central peaks sustain hydrothermal systems longer than simple craters (Fig. 6.6b). In these larger craters, the thermal activity causes the convection of groundwater and meteoric water. There is an extensive zone of fractured rocks on the floor, favorable to the circulation of the prebiotic soup. Moreover, the shock deformation of mineral surfaces stimulates reactions between minerals and active fluids [31, 32]. As we have noted, impact-generated silicate debris in the crater basins would produce clay minerals, and hydrothermal fluid would precipitate pyrites on the crater floor; these minerals could catalyze and polymerize the organic synthesis of prebiotic biopolymers such as nucleic acids and proteins [29, 48, 57, 58, 60].

In a terrestrial setting, the interaction of melt rock with groundwater generates a hydrothermal system (Fig. 6.4). Eventually, both groundwater and rainwater would fill the crater to form a lake, but the crater's floor would have sustained the hydrothermal system for a long time. The crater's rim rises considerably above the lake level and creates an

ideal sequestered basin for the concentration of biomolecules. Impacts on a water-rich planet like Earth or even Mars can generate hydrothermal activity, that is, underwater areas boiling with heat [64]. The high geothermal gradient of the impact-fractured central peak of a complex crater could spew chemicals like the activity at submarine hydrothermal vents and might be an essential source of heat for creating a hydrothermal system. Small complex craters (~5 km in diameter) such as the Gardnos crater of Norway and the Gow crater of Canada would be ideal for prolonged biosynthesis. Evaporative heating and drying of organic compounds near the crater's surface have been proposed as the mechanism for concentrating prebiotic molecules [16, 30, 46, 47].

There is no constraint on the length of time needed for the origin of life. The duration of impact-generated hydrothermal systems is poorly known. In general, the larger the crater, the longer the hydrothermal activity lasts. For example, the hydrothermal system in the 4-km-diameter Kärdla crater in Estonia lasted for several thousand years, whereas the hydrothermal system in the 24-km-diameter Haughton crater in Canada was maintained for more than 10,000 years. We believe that the extremely large Sudbury crater (~250 km) retained hydrothermal activity for ~2 myr [32, 65]. However, during the Eoarchean era, because the crust was relatively thin, the heat flow through it was greater than that on present Earth, and, so, these crater lakes possibly retained hydrothermal activity for more extended periods. Moreover, given the greater abundance of heat-producing isotopes in the early part of Earth's history, the continental crust at that time produced more heat than it does today. Their gradual cooling to form complex craters would have been advantageous for biogenesis, creating both simple and complex organic compounds at different thermal gradients.

6.7 Reconstruction of Impact Crater Lakes in the Vaalbara Supercontinent

The Archean Kaapvaal Craton of South Africa and the Pilbara Craton of Australia are the only two sizeable areas in the world where granite-greenstone terrains ranging in age between 3.6 and 2.7 Ga have been preserved in a relatively pristine state, with rich records of the earliest microfossils. Moreover, these two cratons were once joined together to form an ancient supercontinent. Vaalbara is the name given to the oldest Archean supercontinent, which consisted of the Pilbara and the Kaapvaal Cratons [22]. Remarkably similar lithostratigraphic and chronostratigraphic structural sequences in these two cratons have been noted for the period 3.5–2.7 Ga. Paleomagnetic data from two ultramafic complexes in the cratons showed that at 3.87 Ga, suggesting that these tow cratons were part of the same supercontinent [19]. Both cratons preserve four large meteorite impact events between 3.5 and 3.2 Ga, visible as layers formed when the high temperatures created by impact forces fused the target rocks into small glassy spherules [66]. Occurring in both the Kaapvaal Craton and Pilbara Craton, these spherule layers are the oldest known terrestrial impact products. The spherules resemble the glassy chondrules in the carbonaceous chondrites found in carbon-rich organic compounds of meteorites [67]. Drill core samples of meteoritic components from the Barberton Greenstone Belt of the Kaapvaal Craton confirm the impact origins of these spherule layers [68]. The simultaneous occurrence of impact layers in different regions of the Vaalbara supercontinent suggests large and recurrent impact cratering events during the origin and early evolution of life. Just as they did on the Moon's surface, these impact events must have left numerous scars on the Eoarchean crust of the Vaalbara supercontinent, although evidence for this early cataclysm has long since been obliterated.

We hypothesize that during the Eoarchean and Paleoarchean eras, the surface of the Vaalbara supercontinent looked like the Moon's surface, with thousands of craters ranging in diameter from 1 km to several hundred kilometers [10]. Unlike the Moon, however, the cratered landscape of Vaalbara was filled with water and impact-generated hydrothermal systems, forming ideal incubators for life [20]. These hydrothermal lake sediments contain a rich record of the earliest microbial communities, preserving crucial evidence of the origin of life [52, 61, 69]. Thus, the hydrothermal crater lake environments of the Vaalbara continent hold clues to the locations and likely niches of life's crucibles.

I propose that the Eoarchean and Paleoarchean crust of the Vaalbara supercontinent was heavily pockmarked by meteorite impacts like those on the Moon (Fig. 6.7). As suggested by four distinct impact spherule layers, during a period of light bombardment around 3.5–2.5 Ga, large asteroids struck the Vaalbara continent and blasted out vast crater basins [68, 69]. The second line of evidence of bolide impact in these greenstone belts is an iridium anomaly of meteoritic origin [11]. Iridium is rare in the crust but is abundant in certain carbonaceous asteroids. Chromium isotopes and iridium anomalies confirm the origins of these Archean spherule beds in the multiple impacts of carbonaceous chondrite [70]. Therefore, the large impacts did not stop at ~3.8 Ga but continued throughout the Eoarchean and Paleoarchean eras (~3.8–3.2 Ga). Great floods of molten rock gushed up from the mantle to the granitic crust, filling the crater basins repeatedly with impact-triggered melts and creating ultramafic and basaltic maria or plains associated with hydrothermal lake sediments. Impact-induced ultramafic and mafic volcanic rocks with thin interflows of hydrothermal chert horizons are visible on the Onverwacht Group of the

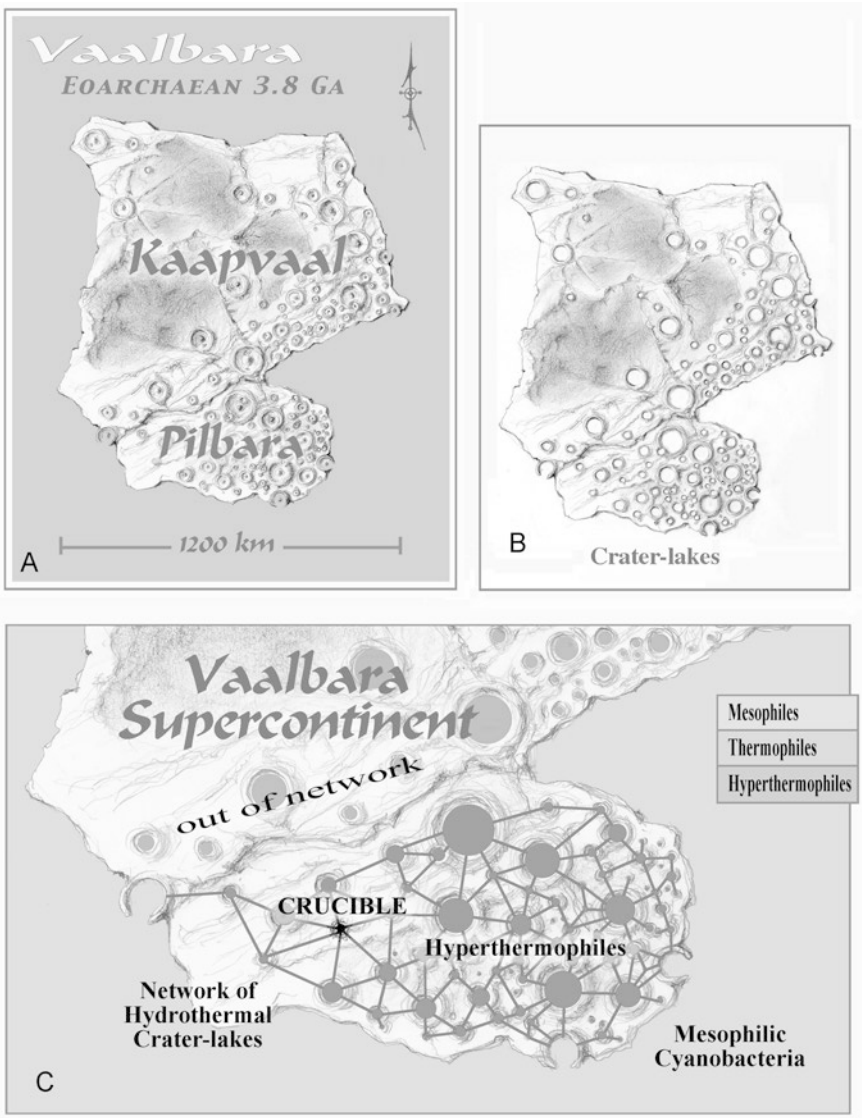


Fig. 6.7 (**a**) Reconstruction of the highly cratered surface of Earth's first supercontinent, Vaalbara, where the granitic crust is pockmarked with innumerable craters like those of the Moon. (**b**) Crater lakes with hydrothermal systems, created by rain and aquifers, rich in cosmic ingredients. (**c**) An underground network formed within the bedrock fractures, interconnecting the various, closely linked crater lakes and exchanging heat and life-building chemical ingredients. These interconnected hydrothermal crater lakes become the ideal cradle for the origin of life

Kaapvaal Craton and on the Warrawoona Group of the Pilbara Craton of the ancient Vaalbara continent [68]. The hydrothermal cherty sedimentary layers are rich in microfossils and have provided critical information about the origin and early evolution of life [52, 69].

Environmental conditions on the Eoarchean surface of the Pilbara Craton and the Kaapvaal Craton were conducive to the origin and diversification of microbial life. These greenstone belts with hydrothermal crater lakes were the ideal location for biosynthesis. Interstellar particles, micrometeorites, small comets, and chondrites were suitable carriers for the safe delivery of cosmic biomolecules to these crater regions. Earth's young atmosphere slowed down these carriers of life's first building blocks, lightly settling as fine dust upon the crater lake surface and enriching them with cosmic ingredients to be mixed by the convection currents of the hydrothermal systems into a rich, prebiotic broth.

While raised rims on the surface separated and isolated these crater lakes, underground cracks and crevices interlinked them, something like the Mono Lake in California, where a series of lakes from higher to lower elevation have a linked flow of groundwater. Instead of a single crucible, then, the closely spaced crater basins on the Vaalbara supercontinent, ranging from 5 to 500 km in diameter, were interconnected through extensive underground networks. Such networks of connected craters had a higher probability than a single crater of forming the ideal crucible systems for the origin of life. Cosmic ingredients and temperature gradients could move from one crater to another through these elaborate underground networks, thus increasing the chances for just the right crucibles in which life took form (Fig. 6.6). The network of crater lakes became increasingly suited to bootstrap prebiotic synthesis and enhanced the conditions for biosynthesis.

6.8 Microbial Colonization in Impact Crater Lakes

The inhabitants of current hydrothermal crater lakes represent relict microbial communities that have remained distinct from other surrounding terrestrial organisms for millions of years, retaining their stamp of extreme antiquity. Today, hydrothermal crater lakes harbor rich ecosystems with variable energy sources, mainly stemming from vents and impact melt rocks. They are favorable habitats for microbial colonization, with many being holdovers from Archean ecosystems that have retained their ancestral characteristics [71, 72]. Thus, the hydrothermal crater lakes of today are a refuge for novel derivatives of the ancient forms of young Earth. The high-temperature vents of modern crater basins are perhaps the oldest ecosystem reminiscent of early Earth. Inside the crater, basins contain reactive gases, dissolved elements, and chemical and thermal gradients ranging on the water surface from hyperthermophilic and thermophilic to mesophilic conditions (Fig. 6.6). The vent microorganisms are self-sufficient. Vent chemicals sustain hyperthermophiles, whereas photoautotrophs requiring solar power enjoy a planktonic lifestyle near the surface. While the fractured rocks of impact craters have been suggested to host deep microbial communities on Earth and potentially other terrestrial planets, direct evidence remains elusive. Still, a few examples of microbial activity in some of these impact craters do suggest that crater lakes are favorable microbial habitats on Earth and perhaps beyond.

The Siljan Crater, Sweden

Formed during the Late Devonian period (~378 Ma) with a diameter of 52 km, the Siljan crater in Sweden is the largest known impact structure in Western Europe. Its drill core samples provide evidence of deep and ancient life. The structure had preserved long-term deep microbial activity that occurred between 80 and 22 million years ago, when conditions, such as temperature, were more favorable than at the time of the impact event. Calcium carbonate and sulfide crystals in the crater's fracture zones have isotopic signatures revealing both microbial methanogenesis and anaerobic oxidation of methane [73].

The Ries Crater, Germany

Formed about 14 Ma, the asteroidal impact that created the Ries Crater (~24 km in diameter) excavated 500–650 m of Triassic–Jurassic and tertiary sedimentary rocks and more than 2 km of the underlying granitic basement. The Ries crater is widely recognized as an analog for Martian craters and specifically ejecta fluidization, postimpact hydrothermal activity, and aqueous sedimentation. The sediments deposited from impact-induced melt rocks such as breccias and glassy melt fragments had a substantial effect on lake water chemistry. In particular, aqueous alteration of the impact glass led to the abundance of zeolite minerals. Rapid weathering ash and melt fragments initially created a highly alkaline crater lake, but the weathering of the felsic crust shifts the pH toward neutral. This observation is similar to the condition of the Gale crater of Mars, which hosted an ancient, potentially habitable lake. Moreover, in both cases, the target rock is granitic [74]. Microbial trace fossils occur in the Ries crater as tubular features in impact glasses. The tubules have complex forms—consistent with the tunneling behavior of microbes—and contain organic molecules associated with biological activity. The Ries crater may have generated hydrothermal activity for as long as 10,000 years, giving microbes ample time to build colonies [75].

The Haughton Crater, Canada

The Miocene Haughton impact structure (~20 km in diameter) on Devon Island in the Canadian High Arctic contain fossils of microbial communities [71]. The site includes massive gypsum-bearing carbonate rocks of the Ordovician age. Microbial communities found in the highly porous and shocked gypsum crystals show two species of cyanobacteria: *Gloeocapsa alpine* and *Nostoc commune*. Microorganisms can better colonize porous minerals and more easily extract nutrients from them. The National Aeronautics and Space Administration (NASA)'s ongoing search for life initiated the Haughton Mars Project, as the freezing environment of the Haughton crater offers a potential Martian analog.

The Lonar Crater, India

Formed entirely within the Deccan Traps around 50,000 years ago, the Lonar Crater of India represents one of the youngest and best-preserved impact structures on Earth. The 1.8-km-diameter simple crater presents an important analog for small craters on the basaltic surfaces of the Moon and Mars (Fig. 6.8). The impact event that generated the Lonar crater probably tapped groundwater supplies in the underlying basaltic aquifer, producing a fluidized ejecta blanket. A hydrothermal system developed as water from this lake began to interact with the hot, porous impact melt deposits. Core samples from the Lonar crater flow show some clay minerals that formed in scorching, hyperthermophilic environments between 130 and 200 °C [76]. Environmental constraints in this highly saline and alkaline soda lake with a microbial community have favored a distinctive and diverse mesophilic microbial community. The Lonar Crater's microbial assemblage includes methylotrophs, anoxygenic purple sulfur and non-sulfur photosynthetic bacteria, and oxygenic cyanobacteria (Fig. 6.7). *Actinobacteria* (24%), *Proteobacteria* (30%), *Firmicutes* (11%), and *Cyanobacteria* (5%) pre-

Fig. 6.8 The Lonar crater is located in the Buldhana District of Maharashtra of western India. It sits inside the Deccan Trap—a massive continental flood basalt volcanism often linked to dinosaur extinction about 66 million years ago. The crater's diameter is approximately 1.8 km, resulting from an impact between 35,000 and 50,000 years ago. This simple crater is around 137 m deep, with an elevated rim roughly 30 m above the surrounding land surface. This warm alkaline lake supports large mesophilic microbial communities. The crater's formation in volcanic basalt provides an excellent analog to impact craters on the Moon. The exposed sloping edge provided condensation reactions in Archean crater lakes. (Courtesy of NASA)

dominate, but other microbes such as *Nitrospirae* (0.41%), *Bacteroidetes* (1.12%), *BD1–5* (0.5%), and *Verrucomicrobia* (0.28%) also occupy the ecosystem [77] (Fig. 6.8).

6.9 The Plausible Cradle for Life's Beginnings

Four billion years ago, simple organic compounds assembled together to form the first life on young Earth. Being both microscopic and hyperthermophilic, these more complex combinations could grow and reproduce. However, at the time of their emergence, Earth was completely devoid of what we would recognize today as a suitable environment for living things. Instead, the hostile and extreme environment of the young planet sustained widespread impact and volcanic activity. Where on Earth could life-forming processes find enough shelter to proceed?

In the quest for the cradle of life, the essential first step is to determine the precise environment where the cosmic ingredients and other building blocks of life could be concentrated and synthesized into more complex molecules and where energy for protometabolism would be available. Moreover, that environment should be geologically compatible with a planet lacking any plate tectonic activity. The lifestyle of living hyperthermophiles, the most primitive organisms known, provides an important clue to the cradle's paleoenvironment: an anoxic hydrothermal vent environment that supplies rich minerals and energy [24, 45]. Indeed, hydrothermal conditions are environments conducive to nascent life. Moreover, genetic analyses lend a certain amount of corroboration for the emergence of life in hydrothermal systems, for these studies suggest that LUCA lived in a hydrothermal environment. LUCA's closest living relatives are *Clostridium* bacteria and methanogenic archaea, gaining nutrients from H_2S, H_2, CO_2, transition metals, and sulfur [26, 78].

Currently, the top four hydrothermal locations for abiogenesis are submarine vents such as black smokers and Lost City and terrestrial vents such as 'Darwin's warm little ponds' and impact crater lakes. While each environment has pros and cons, in recent times, the long-held view that life originated in a marine environment has been contested. Many now consider that oceanic sites will dilute rather than concentrate organic molecules for biosynthesis. Moreover, the lack of plate tectonics during the early Archean has cast doubt about the very presence of hydrothermal vents on young Earth. Moreover, primitive cellular membranes appear to assemble more easily in freshwater than in saltwater. Moreover, the cytoplasm of the living cells more resembles the chemical nature of terrestrial vents. Finally, the earliest biosignatures of microbial life are not found in the deep sea but in the terrestrial environment.

Regarding the two terrestrial alternatives, both hot springs and impact craters are attractive models. However, hot springs are relatively short-lived and were probably less frequent on the granitic crust. Both surface sites do have access to light energy, chemical energy, and concentrating mechanisms. However, in the Eoarchean, the only

land exposed above the sea level was the granitic crust, where impact craters were likely more available than volcanic hot springs during the Heavy Bombardment period. The floors of impact crater lakes also had pyrite and clay surfaces for polymerization, dissolved gases like hydrogen and methane, solutes such as ferrous iron, energy like ATP for protometabolism, and long residence times for hydrothermal activity.

Impact-generated hydrothermal systems may have provided the most favorable environments for the prebiotic synthesis of cosmic ingredients and life's origin. The impact cratering record on Earth during the Eoarchean suggests that hydrothermal crater lakes were ubiquitous after asteroidal impacts on water-rich crustal surfaces. Making water, heat, dissolved chemicals, and nutrients available for extended periods, these hydrothermal systems were prime locations for abiogenesis and colonization by hyperthermophilic microorganisms. They could also have provided a sanctuary for hyperthermophiles during the Late Heavy Bombardment. Mineral surfaces on the crater floor could have helped catalyze complex chemical reactions, polymerization, and available energy for primitive metabolism. In our view, the interconnected impact crater lakes with a wide range of temperatures, pH values, and concentrations of cosmic ingredients had the right conditions for the synthesis of membranes, peptides, RNAs, viruses, DNAs, and the first cells. The oldest fossils from the Vaalbara supercontinent are consistent with the hydrothermal crater environment for early life [66, 67]. Impact cratering hydrothermal systems appear to be common on young Earth and its rocky neighbor in the solar system. Especially in planetary bodies such as Mars, which are otherwise geologically dead, the fact that such systems could have provided rare havens for life has considerable astrobiological implications [32]. The crater lakes are the most likely to be the long-sought cradle of life. Life probably got started on the surface of Earth, such as around impact craters, where it found sequestered hydrothermal basins offering refuge during a bombardment. When the cratering activity diminished, life could have then spread to the tranquil ocean surfaces, eventually harnessing the Sun's energy and distributing globally [20].

6.10 Conclusions

The greenstone facies of the Eoarchean crust, though limited in distribution in the oldest cratons, constitute the most valuable archive for unraveling the mystery of life. Life appears to have originated near the end of the most violent period of Earth's history, the Late Heavy Bombardment between 4.1 and 3.8 billion years ago. Plate tectonics has recycled and resurfaced the primitive crust of our planet and erased the evidence of this early onslaught. Nevertheless, ancient craters on the surfaces of the Moon and Mercury show just how heavy the meteorite showers once were. Before the onset of plate tectonics, impact cratering was the primary geological force and likely played an important role in the origin of life. Meteorite impacts not only delivered critical building blocks of life to young Earth but they also happened to create thousands of impact crater lakes with hydrothermal systems on the Eoarchean continental crust. Relative to modern impact flux, asteroidal impacts were larger and more numerous during the Late Heavy Bombardment. Many of these craters were closely spaced and interconnected by underground fracture systems, capable of exchanging heat and life-building molecules. The heat, convection currents, and spewing gases of sequestered crater lake environments concentrated the cosmic building blocks of life, fueling many of the chemical reactions necessary for abiogenesis. These hydrothermal crater lakes became the perfect cradles for prebiotic synthesis.

References

1. Mojzsis SJ, Harrison TM, Pidgeon TT (2001) Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. Nature 409:178–18115
2. Harrison TM, Schmitt AK, McCulloch MT, Lovera OM (2008) Early (>4.5 Ga) formation of terrestrial crust: Lu-Hf, ∂^{18}O, and Ti thermometry results for Hadean zircons. Earth Planet Sci Lett 268:476–486
3. Miller SR, Mueller PA, Meert JG, Kamenov GD, Pivarunas AF, Sinha AK, Pandit MK (2018) Detrital zircons reveal evidence of Hadean crust in the Singhbhum Craton, India. J Geol 126:541–552
4. Tang M, Chen K, Rudnick RL (2016) Archean upper crust transition from mafic to felsic marks the onset of plate tectonics. Science 351:373–375
5. Hawkesworth CJ, Cawood PA, Dhuime B (2016) Tectonics and crustal evolution. GSA Today 26:4–11
6. Mojzsis SJ, Harrison TM (2000) Vestiges of a beginning. Clues to the emergent biosphere recorded in the oldest known sedimentary rocks. GSA Today 10:1–5
7. Bowring SA, Williams IS (1999) Priscoan (4.00–4.03) orthogenesis from northwestern Canada. Contrib Mineral Petrol 134:13–16
8. Stern RJ (2005) Evidence from ophiolites, blueschists, and ultrahigh-pressure metamorphic terranes that the modern episode of subduction tectonics began in Neoproterozoic time. Geology 33:5557–5560
9. O'Neil J, Carson RW (2017) Building Archean Cratons from Hadean mafic crust. Science 355:1199–1202
10. Marchi S, Bottke WF, Elkins-Tanton LT, Bierhaus M, Wuennemann K, Morbidelli A, Kring DA (2014) Widespread mixing and burial of Earth's Hadean crust by asteroid impacts. Nature 511:578–582
11. Glikson AY (2010) Archean asteroid impacts, banded iron formations and MIF-S anomalies. Icarus 207:39–44
12. Grieve RAF, Cintala MJ, Therriault AM (2006) Large-scale impacts and the evolution of Earth's crust: the early years. In: Reimold WU, Gibson RL (eds) Processes of the early earth, Geological Society of America special papers, vol 405, pp 23–31
13. Kring DA (2000) Impact events and their effect on the origin, evolution, and distribution of life. GSA Today 10(8):1–7
14. Melosh HJ (1989) Impact cratering: a geologic process. Oxford University Press, New York

15. Garde AA, McDonald I, Dyck B, Keulen N (2012) Searching for giant, ancient impact structures on Earth: the Mesoarchaean Maniitsoq structure, West Greenland. Earth Planet Sci Lett 337:197–210
16. Cockell C, Lee P (2002) The biology of impact craters – a review. Biol Rev 77:279–310
17. Hansen VL (2015) Impact origins of Archean Cratons. Lithosphere 7. https://doi.org/10.1130/L371.1
18. Hansen VL (2007) Subduction on early Earth. Geology 35:1059–1062
19. Zegers TE, Ocampo A (2003) Vaalbara and tectonic effects of a mega impact in the early Archean 3470 Ma. In: Abstract, third international conference on large meteorite impacts, Nordington
20. Chatterjee S (2018) The hydrothermal impact crater lakes: the crucibles of life's origin. In: Kolb VM (ed) Handbook of astrobiology. CRC Press/Taylor & Francis, Boca Raton, pp 265–295
21. De Wit M, Ashwal LD (1997) Greenstone belts. Clarendon Press, Gloucestershire
22. Zegers TE, De Wit MJ, Dann J, White SH (1998) Vaalbara, Earth's oldest assembled continent? A combined structural, geochronological, and palaeomagnetic test. Terra Nova 10:250–259
23. McCollom TM, Shock EL (1997) Geochemical constraints on chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal systems. Geochim Cosmochim Acta 61:4375–4439
24. Stetter KO (2006) Hyperthermophiles in the history of life. Philos Trans R Soc 361:1837–1843
25. Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposals for the domains Archaea, Bacteria, and Eukarya. Proc Natl Acad Sci U S A 87:4576–4579
26. Akanuma S, Nakajima Y, Yokobori S, Kimura M, Nemoto N, Mase T, Miyazono K, Tanokura M, Yamagishi A (2013) Experimental evidence for the thermophilicity of ancestral life. Proc Natl Acad Sci U S A 110:11067–11072
27. Baross JA, Hoffman SE (1985) Submarine hydrothermal vents and associated gradient environment as sites for the origin and evolution of life. Orig Life 15:327–345
28. Martin W, Russell MJ (2007) On the origin of biochemistry at an alkaline hydrothermal vent. Philos Trans R Soc B362:1887–1926
29. Wachterhäuser G (1993) The cradle of chemistry of life: on the origin of natural products in a pyrite-pulled chemoautotrophic origin of life. Pure Appl Chem 65:1343–1348
30. Deamer DW (2011) First life: discovering the connections between stars, cells, and how life began. University of California Press, Berkeley
31. Cockell CS (2006) The origin and emergence of life under impact bombardment. Philos Trans R Soc B361:1845–1856
32. Osinski GR, Tornabene LL, Banerjee NR, Cockell CS, Flemming R, Izawa MRM, McCutcheon J, Parnell J, Preston LJ, Pickersgill AE, Pontefract A, Sapers HM, Southam G (2013) Impact-generated hydrothermal systems on Earth and Mars. Icarus 224:347–363
33. Dodd MS, Papineau D, Grenne T, Slack JF, Rittner M, Prajno F, O'Neil JO, Little CTS (2017) Evidence for early life in Earth's oldest hydrothermal vent precipitates. Nature 543:60–65
34. Furnes H, Banerjee NR, Muehlenbachs K, Staudigel H, de Wit M (2004) Early life recorded in Archean pillow lavas. Science 304:578–581
35. Hofmann A (2011) Archean hydrothermal systems in the Barberton Greenstone belt and their significance as a habitat for early life. In: Golding SD, Glikson M (eds) Earliest life on Earth: habitats, environments and methods of detection. Springer, New York, pp 51–78
36. Holm NG (1972) Why are hydrothermal vent systems proposed as plausible environments for the origin of life? Orig Life Evol Biosph 22:5–14
37. Nisbet EG, Sleep NH (2001) The habitat and nature of early life. Nature 409:1083–1091
38. Von Damm KL (1990) Seafloor hydrothermal activity: black smoker chemistry and chimneys. Annu Rev Earth Planet Sci 18:173–204
39. Kelley DS, Baross JA, Delaney JR (2002) Volcanoes, fluids, and life at mid-ocean ridge spreading centers. Annu Rev Earth Planet Sci 30:385–491
40. Miller SL (1953) A production of amino acids under possible primitive Earth conditions. Science 117:528–529
41. Alberts SV, Van de Goosenberg JLCM, Driessen AJM, Konings WN (2000) Adaptations of the archaeal cell membrane to heat stress. Front Biosci 5:813–820
42. Sousa F, Thiegart T, Landan G, Nelson-Sathi S, Pereira IAC, Allen JF, Lane N, Martin WF (2013) Early bioenergetic evolution. Philos Trans R Soc B 368:2013988. https://doi.org/10.1098/rstb.2013.0088
43. Kelley DS, Karson JA, Fruh-Green GL et al (2005) A serpentine-hosted ecosystem: the Lost City hydrothermal field. Science 307:1428–1434
44. Martin W, Baross J, Kelley D, Russell MJ (2008) Hydrothermal vents and the origin of life. Nat Rev Microbiol 6:805–814
45. Lane N, Martin WF (2012) The origin of membrane bioenergetics. Cell 151:1406–1416
46. Deamer DW (2019) Assembling life: how can life begin on Earth and other habitable planets. Oxford University Press, New York
47. Damer B, Deamer D (2015) Coupled phases and combinatorial selection in fluctuating hydrothermal pools: a scenario to guide experimental approaches to the origin of cellular life. Life 5:872–887. https://doi.org/10.3390/life5010872
48. Hazen RM (2005) Genesis: the scientific quest for life. Joseph Henry Press, Washington, DC
49. Mulkidjanian AY, Bychkov AB, Diprova DV, Galperin MY, Koonin EB (2012) Origin of first cells at terrestrial, anoxic geothermal fields. Proc Natl Acad Sci U S A 109:E821–E830
50. Pearce BKD, Pudritz RE, Semenov DA, Henning TK (2017) Origin of the RNA world: the fate of nucleobases in warm little ponds. Proc Natl Acad Sci 114:11327–11332
51. Bernhardt HS, Tate WP (2012) Primordial soup or vinaigrette: did the RNA world evolve at acidic pH? Biol Direct 7:4. www.biology--direct.com/content/7/1/4
52. Djokic T, Van Kranendonk MJ, Campbell KA, Walter MR, Ward CR (2017) Earliest signs of life on land preserved in ca, 3.5 Ga hot spring deposits. Nat Commun. https://doi.org/10.1038/ncomms15263
53. Clark BC, Kolb VM (2018) Comet Pond II: synergistic intersection of concentrated extraterrestrial materials and planetary environments to form procreative Darwinian ponds. Life 8:12. https://doi.org/10.3390/life80200012
54. Kring D, Cohen BA (2002) Cataclysmic bombardment throughout the inner Solar System 3.9-4.0 Ga. J Geophys Res Planets 107:E2. https://doi.org/10.1029/2001JE001529
55. Bada JL, Korenga J (2018) Exposed areas above sea level on Earth >3.5 Gyr ago: implications for prebiotic and primitive biotic chemistry. Life 8:55. https://doi.org/10.3390/life8040055
56. Lewis RE, Young HW (1982) Geothermal resources in the Banbury Hot Springs area, Twin Falls county, Idaho. US Geol Surv Water-Suppl Pap 2186:1–27
57. Chatterjee S (2016) A symbiotic view of the origin of life at hydrothermal impact crater lakes. Phys Chem Chem Phys 18:20033–20046
58. Greshake A, Kloek W, Arndt P, Maetz M, Flynn GJ, Bajt S, Bischoff A (1988) Heating experiments stimulating atmospheric entry heating of micrometeorites: clues to their parent body source. Meteorit Planet Sci 33:267–290
59. Chatterjee S, Yadav S (2019) The origin of prebiotic information system in the peptide/RNA world: a simulation model of the evo-

lution of translation and the genetic code. Life 9:25. https://doi.org/10.3390/life9010025
60. Huber C, Wachterhäuser G (1998) Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces: implications for origins of life. Science 281:670–672
61. Wacey D, Kilburn MS, Saunders M, Cliff J, Brasier MD (2011) Microfossils of sulfur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. Nat Geosci 4:698–702
62. Martin WF, Sousa FL, Lane N (2014) Energy at life's origin. Science 344:1092–1093
63. Grieve RAF (1990) Impact cratering on Earth. Sci Am 261:66–73
64. Farmer JD (2000) Hydrothermal systems: doorways to early biosphere evolution. GSA Today 10:1–9
65. Abramov O, Kring DA (2004) Numerical modeling of an impact-induced hydrothermal system at the Sudbury crater. J Geophys Res 109. https://doi.org/10.1029/2003JE002213
66. Beyerly GR, Lowe DL, Wooden JL, Xie X (2002) An Archean impact layer from the Pilbara and Kaapvaal Cratons. Science 297:125–1327
67. Lowe DR, Byerly GR, Kyte FT, Shukolyukov A, Asaro F, Krull A (2003) Geological and geochemical record of 3400-million-year-old terrestrial meteorite impacts. Science 245:959–962
68. Koeberl C, Schulz T, Reimold WU (2015) Remnants of Early Archean impact deposits on Earth: search for a meteorite component in the BARBS and CT3 drill cores (Barberton Greenstone Belt, South Africa). Process Eng 103:310–317
69. Westall F, de Wit MJ, Dann J, van der Gaast S, Ronde CEJ, Gerneke D (2001) Early Archean fossil bacteria and biofilms in hydrothermally influenced sediments from the Barberton Greenstone belt, South Africa. Precambrian Res 106:93–116
70. Kyte FT, Shukolyukov A, Lugmair GW, Lowe DR, Byerly GR (2003) Early Archean spherule beds: chromium isotopes confirm origin through multiple impacts of projectiles of carbonaceous chondrite type. Geology 31:283–286
71. Parnell J, Lee P, Cockell CS, Osinski GR (2004) Microbial colonization in impact-generated hydrothermal sulphate deposits, Haughton impact structure, and implications for sulfate on Mars. Int J Astrobiol 3:247–225
72. Reysenbach AL, Cady SL (2001) Microbiology of ancient and modern hydrothermal systems. Trends Microbiol 9:79–86
73. Drake H, Roberts NMW, Heim C, Whitehouse ML, Silijestorm S, Koojiman E, Broman C, Ivarsson M, Astrom ME (2019) Timing and origin of natural gas accumulation in the Siljan impact structure, Sweden. Nat Commun 10:4736. https://doi.org/10.1038/s41467-019-17728-y
74. Stuekan EE, Tino C, Arp G, Jung D, Lyons TW (2020) Nitrogen isotope ratios trace high pH conditions in a terrestrial Mars analog site. Sci Adv 6:eaay3440
75. Sapers HM, Osinski GR, Banerjee NR, Preston LJ (2015) Enigmatic tubular features in impact glass. Geology 43:635–638
76. Hagerty JJ, Newsom HE (2003) Hydrothermal alteration at the Lonar Lake impact structure, India: implications for impact cratering on Mars. Meteorit Planet Sci 38:365–381
77. Paul D, Kumbhare SV, Mhatre SS, Chowdhury SP, Shett SA, Marathe NP, Bhute S, Shouche YS (2016) Exploration of microbial diversity and community structure of Lonar lake: the only hypersaline meteorite crater-lake within basalt rock. Front Microbiol. https://doi.org/10.3389/fmicb.2015.01553
78. Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF (2016) The physiology and habitat of the last common ancestor. Nat Microbiol 1:161116. https://doi.org/10.1038/nmicrobiol.2016.116

Bioenergetics and Primitive Metabolism

7

Life began when a system of encapsulated polymers captured energy and nutrients from the environment and then used that energy to grow and reproduce.

—David Deamer, 2011

Bioenergetics is the study of the flow of energy through a living system. The origin of life requires a source of energy. Energy (capacity to perform work or supply heat) is generated when a force operates on an object over a distance. There are two types of energy—potential and kinetic. Potential (or stored energy) is the energy of state or position, such as the energy stored in chemical bonds. Kinetic energy is the energy of motion; it includes electric energy, light, and heat. Molecules have kinetic energy because they are constantly in motion. When energy moves from one location to another, it is called energy transfer. In prebiotic synthesis, it is more beneficial to consider energy as having the capacity for change. Mixtures of chemical compounds in the vent environment can contain potential energy stored in the electronic structure of chemical bonds. However, energy can also be added to certain compounds, which can then undergo a chemical reaction that would otherwise not occur. Our senses can detect the energy conveyed by sunlight, winds, and tides, but most forms of energy are invisible, such as the energy contained by fossil fuels and nutrients.

7.1 Thermodynamics and Energy Conversions

Thermodynamics studies the relationship between work and energy. It describes the energy changes that occur during a chemical reaction. Two thermodynamic laws govern energy transformation. The first law asserts that energy is neither created nor destroyed; it can only be converted from one form to another. All energy comes from somewhere and then goes somewhere else. All living systems obey this law. In every spontaneous chemical reaction, there are two quantifiable changes in energy. The 'enthalpy change' is the measurement of energy in heat given off by a reaction.

The second law of thermodynamics states that chemical reactions spontaneously increase the 'entropy' or disorder of an isolated system. Although energy cannot be created or destroyed, when it is converted from one form to another, some of it becomes unavailable. This unusable energy is lost in disorder. The total energy of a biological system is called enthalpy (H), the usable energy is called Gibbs free energy (G), the unusable energy is entropy (S), and absolute temperature is given in Kelvin units (T).

These relationships can be written in a simple equation:

$$H = G + TS$$

Because we are interested in usable energy, we rearrange the expression:

$$G = H - TS$$

Although we cannot measure G, H, or S absolutely, we can measure the change at a constant temperature, where the Greek letter delta (Δ) means difference or change:

$$\Delta G = \Delta H - T\Delta S$$

ΔG is the change in free energy, ΔH is the change in enthalpy, and ΔS is the change in entropy. For example, our bodily temperature remains above room temperature because of the heat energy (ΔH) given off by ATP's hydrolysis reaction. Since cellular reactions occur under normal pressure and temperature conditions, the change in enthalpy is about the same as the change in total energy during the reaction. Moreover, since free energy change (ΔG) is the amount of energy in a system available to perform helpful work at constant temperature and pressure, the change in entropy (ΔS) is a measure of the proportion of the total energy change that the system cannot use in performing work. For example, when we dissolve a salt crystal, it

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_7

releases sodium (Na^+) and chloride (Cl^-) ions. The sodium and chloride atoms that compose salt are highly organized in the crystal but become disorganized when they dissolve in the water, thus increasing entropy.

The equation $\Delta G = \Delta H - T\Delta S$ tells us whether free energy is released ($\Delta G \leq 0$) or consumed ($\Delta G \geq 0$) in a chemical reaction. It follows from the equation that a reaction with a large positive change in entropy (ΔS) will generally tend to have a negative ΔG and therefore occur 'spontaneously.' A chemical reaction is spontaneous if the changes in enthalpy and entropy are both favorable. In other words, when a reaction occurs spontaneously, the change in free energy is negative. These reactions, which release free energy, are called 'exergonic' (or favorable) and have a negative ΔG. Conversely, a decrease in entropy will tend to make ΔG more positive. When a reaction is nonspontaneous and requires energy input, the change in free energy is positive. Reactions that proceed toward completion only with the addition of free energy from the environment are 'endergonic' (or unfavorable) and have a positive ΔG. If the necessary free energy is not available, then the reaction does not occur.

Spontaneous reactions could have happened on their own without requiring an input of energy from the protocells. According to the second law of thermodynamics, reactions can be spontaneous only if they increase entropy. Dissolving salt in water or burning a candle are examples of spontaneous reactions with an increase of disorder. Because the spontaneous reaction may break down an ordered reactant, such as a glucose molecule, into smaller, more disordered products such as carbon dioxide and water, such exergonic reactions offer free energy, and other molecules take it up. In a nonspontaneous reaction, however, additional input of energy is required to do the job.

The nonspontaneous reaction may make a single product, such as a protein (a highly ordered polymer), out of many smaller reactants, such as amino acids (less ordered). Such endergonic reactions require energy to occur. Protocells maintain their order by constantly using their stored energy to perform nonspontaneous reactions that decrease entropy, and, to do this, they require energy input. Even exergonic reactions need a small amount of energy input to get going before proceeding with their energy-releasing steps. This small amount of energy input, necessary to all chemical reactions, is called the 'activation energy.'

Finally, in thermodynamics, there are two kinds of systems: closed and open. A closed system cannot transfer energy to its surroundings. Any reaction in a closed system ultimately reaches equilibrium in which no further energy changes can occur because all the free energy has been spent, and so entropy has reached a maximum. In contrast, an open system can exchange both energy and matter with its surroundings. For instance, organisms are open systems that exchange energy with their environments as they draw energy-storing molecules and release energy to their surroundings by doing work. Living systems can clearly achieve a local reduction in their entropy as they grow and develop by creating structures of greater internal energy out of the nutrients they absorb.

7.2 Energy Sources for Emerging Life

All living organisms require energy to grow, reproduce, maintain their structures, and respond to their environments. Before life began on early Earth, various sources of energy were available, with the primary source being the Sun itself. However, the earliest hyperthermophilic life powered itself through reactions involving hydrogen gas, sulfur, or ammonia, available in vent environments. Organisms generally acquire energy from either chemical oxidation or light. Autotrophs receive their useful free energy from sources other than food: either from oxidative chemical reactions (chemoautotrophs) or from sunlight (photoautotrophs). Chemoautotrophs create their own energy and biological materials from inorganic chemicals (ammonia, methane, or hydrogen sulfide). Photoautotrophs capture energy from sunlight to make their own food. We believe that the first cells were anaerobic chemoautotrophs and that photosynthesis evolved only much later. How did the protocells, during prebiotic synthesis in hydrothermal vents, harness energy?

There were plenty of energy sources on prebiotic Earth, in the form of sunlight, ultraviolet radiation, electrical discharges, shock waves, heat, and chemicals from vent environments. Some of the energy sources were global; others were in localized vent environments. Emerging life likely used a variety of the available energy sources to produce a more complex set of polymers from relatively simple molecules. Deamer [1, 2] identified four levels of energy resources that might have been available on prebiotic Earth to drive primitive metabolism and synthesis of polymers. The first level included photochemical energy available in ultraviolet light, atmospheric electrical discharge, and geochemical energy. The energy level of photochemical energy is relatively high and is modeled in the laboratory by electrical discharge and ultraviolet light. The second level of energy would have come from concentrating dilute solutions of cosmic ingredients in the vent environment and concentrating monomers by evaporation. The third level used a series of low-energy reactions that incorporated condensation processes by which monomers were assembled into random polymers on the mineral surfaces. The final level was the energy flow in protometabolic networks. Protocells would have captured the local environment's energy and then transferred it to energy carriers like ATP (adenosine triphosphate). ATP is the primary energy currency for organisms. The

metabolic and catabolic processes in emerging life synthesized ATP from materials available in the vent environment and broke it down to adenosine diphosphate (ADP) through organic phosphate catabolic reactions.

7.3 Global Energy Sources

In prebiotic Earth, several energy fluxes were available globally, such as solar energy, lightning, and impact shock waves that could have driven chemical evolution, leading to increasingly complex organic molecules. Although external to the planet, ultraviolet solar radiation, solar wind, cosmic rays, and shock waves had a global distribution.

7.3.1 Solar Energy

Just as it is today, radiation from the Sun was the primary energy source on our planet before life began. In this oxygen-poor prebiotic world, solar energy may have provided the 'jolt' necessary to transform simple organic molecules into more complex ones. Because of the unique characteristics of photochemistry, solar energy can affect a larger variety of chemistries than energy derived from other sources. Solar UV radiation could have driven the synthesis of small reactive molecules, such as formaldehyde and hydrogen cyanide (HCN), which would have had the chemical potential to produce more complex molecules such as amino acids, sugars, purine, and pyrimidine bases, lipid-like amphiphiles, and carbohydrates [2, 3]. Solar energy helped periodic evaporation of terrestrial hydrothermal systems for the concentration of biomolecules.

7.3.2 Impact Shock Waves

The lunar impact record suggests a high bolide influx until 3.8 billion years ago. As we have discussed, asteroids and comets may have been responsible for delivering intact organic molecules and volatiles to early Earth. Additionally, the energy released by shock waves from impacting meteorites could have synthesized organic compounds such as hydrogen cyanide and amino acids from atmospheric heating and the postimpact recombination of simple organic components [4]. Experiments with high-energy lasers have demonstrated how repeated shocks to a methane-rich mixture can produce hydrogen cyanide, acetylene, and even amino acids. [5]. Such laboratory simulations suggest that organics might have been generated by shock processing of the atmosphere or during the impact itself. The production of atmospheric HCN could result in the generation of ferrocyanide and phosphate salts (containing important ingredients for the origins of metabolism), which then rain down on the surface and are concentrated through evaporation [6].

7.4 Energy Sources in Localized Environments

It is reasonable to assume that the prebiotic environment's energy drove the synthesis of organic compounds required for life's origin. Prebiotic Earth's hydrothermal vent environment was far from equilibrium, so simple chemical reactions were occurring simultaneously. Various energy sources were available for prebiotic synthesis. In hydrothermal impact crater lakes with a flow of hot fluids beneath and up to the surface, heat comes from the underlying magma or the hot water generated by convection currents. Other than thermal energy and electrochemical energy, another source of potential chemical energy available in organic compounds delivered by comets and asteroids was reduced carbon compounds that could undergo chemical modifications in metastable states to organic compounds if exposed to minerals surfaces in hydrothermal settings [7].

7.4.1 Thermal Energy

The cosmic ingredients in the hydrothermal crater lakes would be present in highly diluted solutions. The abundance of heat in the hydrothermal vent environment would have accelerated the chemical reactions needed to provide the foundations of life. Additionally, the large amount of energy introduced by evaporation would have caused a variety of monomers such as amino acids and nucleotides exposed to wet-and-dry cycles to polymerize in a condensation reaction, linked by peptide bonds and phosphodiester bonds, respectively. Thermal energy from the convection current generated by the volcanic heat inside the crater basin could have also promoted the mixing of cosmic ingredients and concentrations of monomers (Fig. 6.4). Dissipative structures produced by the convection current would have maintained the far-from-equilibrium conditions that sustained protocells and set off chemical reactions involving the ingredients from asteroids. The moderate heat of the vent could either speed up or drive a chemical reaction that would not be able to occur otherwise.

7.4.2 Chemical Energy

Prebiotic protocells might have used chemical energy as well. The most likely source of chemical energy is glyceraldehyde, a chemical that could also act as an autocatalyst for prebiotic condensation reactions [8]. Glyceraldehyde

was probably synthesized from formaldehyde in atmospheric reactions. The chemical resemblance of this model to glycolysis in living cells gives it the potential to develop a biological metabolism. Another form of prebiotic chemical energy includes chemically activated monomers such as amino acids (activated form: thio-amino acids), nucleotides (activated forms: di- and triphosphate), and fatty acids (activated forms: anhydrides, thioesters) [2, 9, 10]. Sugar redox disproportionation is the major source of free energy for the activation of these monomers. These activated monomers help polymerize nucleic acids and proteins and develop lipid bilayers in energetically favorable (downhill) chemical reactions.

Phosphate esters and anhydrite dominate the living world. The sugar-phosphate backbone on nucleic acids is negatively charged with hydrogen ion. So, when mRNA is synthesized, an ATP, the primary instant energy source released through hydrolysis of its terminal phosphate group is used to link the nucleotides of mRNA together in a chain. Modern metabolism would not work without ATP and phosphate, but, in primordial biochemical networks, energy currencies must have been simpler. Phosphate is an essential component in ATP, and the abiotic formation of phosphorylated metabolism is a central issue in prebiotic chemistry. In virtually every metabolic pathway step, one or more compounds must be phosphorylated to activate the process. The bond energy of pyrophosphate ($P_2O_7^4$), which has the same chemical energy content as the second and third phosphate groups in ATP, could have been a significant source of chemical energy in the reactions leading to the origin of life [11]. Pyrophosphate bonds readily form when inorganic phosphate is heated and dried to temperatures in the range of 160 to 200 °C, and conditions are easily met in the hydrothermal vent. Phosphorylation reactions must have been assimilated in primitive metabolic pathways, and the frequent meteorite bombardment of young Earth most likely delivered the reactive phosphorous that was incorporated into prebiotic molecules when released in water, thus making ATP available in the hydrothermal vent environment [12].

7.4.3 Electrochemical Energy

Oxidation/reduction (redox) reactions are crucial to supporting chemosynthesis. Molecules that gain electrons are 'reduced'; those that gain them are 'oxidized.' Vent fluid transports various chemically reduced compounds, including hydrogen, sulfide, methane, iron, and manganese, in variable proportions from the deep crust to the crater floor. Today, various hyperthermophiles living in hydrothermal vent environments use these inorganics and gases as sources of electrons. The protocells could have captured energy by transporting electrons from donors like hydrogen, carbon, or hydrogen sulfide to electron acceptors such as oxygen, hydrogen sulfide, or iron, which were present either in a solution or on a mineral surface. Chemical entities carried by the rising water of the vent from the basin's bottom were not in thermodynamic equilibrium. Thus, protocells could have easily harnessed the energy of the naturally occurring chemical reactions. Protocells would have employed these various processes, collectively called chemosynthesis, to harness energy in the hydrothermal vents. Here are some examples of these redox reactions in hydrothermal vents:

Hydrogen oxidation

$$2H_2\,(\text{hydrogen}) + O_2\,(\text{oxygen}) \rightarrow 2H_2O(\text{water}) + \text{energy}$$

Pyrrhotite oxidation

$$FeS(\text{pyrrhotite}) + H_2S(\text{hydrogen sulfide}) \rightarrow FeS_2\,(\text{pyrite}) + H_2\,(\text{hydrogen}) + \text{energy}$$

Sulfur reduction

$$S(\text{sulfur}) + H_2\,(\text{hydrogen}) \rightarrow H_2S(\text{hydrogen sulfide}) + \text{energy}$$

Methane production

$$CO_2\,(\text{carbon dioxide}) + 4H_2\,(\text{hydrogen}) \rightarrow CH_4\,(\text{methane}) + 2H_2O(\text{water}) + \text{energy}$$

7.5 Metabolic Energy

The term 'metabolism' describes the complex network of reactions that enable organisms to generate energy and assemble the molecules they need to live, grow, and reproduce. Metabolism follows metabolic 'pathways,' the 'flow' of chemical reactions, each catalyzed by a series of enzymes and each acting on the previous enzyme's product. Along a metabolic pathway, an initial molecule goes through a series of intermediate states on the way to a final product. The successions of reactions can be reliably repeated, with the same precursor entering and the same end product leaving in each cycle. Metabolic pathways connect with each other to form a complex interlocking web. Metabolism has two opposite processes: 'anabolic' pathways synthesize macromolecules such as glucose, whereas 'catabolic' pathways degrade macromolecules

and take their usable free energy. The first requires energy, and the second produces energy:

$$\text{Anabolic reaction}: 6CO_2 + 6H_2O + \text{energy} \rightarrow C_6H_{12}O_6\,(\text{glucose}) + 6O_2$$

$$\text{Catabolic reaction}: C_6H_{12}O_6\,(\text{glucose}) + 6O_2 \rightarrow 6CO_2 + 6H_2O + \text{energy}$$

For instance, protein enzymes are extremely important for metabolic functioning. By lowering the activation energy, they speed up biochemical reactions so that metabolism occurs quickly. However, before the availability of such enzymes, iron–sulfur (FeS) minerals could have served as catalysts at life's origin. The enzymes that catalyze these reactions today contain small iron, nickel, and sulfur clusters at their cores, with a structure highly similar to those found in vents. Iron and nickel may have come from iron meteorites. Each kind of enzyme selects its substrate from the metabolic pool. Enzymes change shape when they bind with the substrate.

The Krebs cycle, also called the tricarboxylic acid cycle (TCA) cycle, is at the heart of the metabolic network. It consists of a circular chain of reactions that generate precursors of amino acids and lipids, which are then used to build proteins, membranes, and molecules that help a cell produce its energy. Considered to be the biochemical currency of life, adenosine triphosphate (ATP) provides the energy to drive the many processes of living cells.

7.5.1 The Origin of Metabolism

Modern metabolism would not work without ATP and phosphate, but energy currencies might have been more straightforward in the prebiotic world. The development of metabolism was a crucial part of prebiotic synthesis, but its origin in the prebiotic vent environment before the advent of the enzyme is unclear. Researchers have proposed two different views regarding the stage at which metabolism originated. Oparin hypothesized that metabolism arose in the primordial soup during abiogenesis and preceded genetic replication [13]. Alternatively, Haldane claimed that metabolism arrived after the emergence of genes [14]. However, metabolism is not always a genetically directed process and can develop nonenzymatically. The potential for early metabolic pathways in vents supports the idea that basic metabolisms came first and were then later encoded into genetic material [10]—the concept of 'protometabolism' bridges the gene–metabolism divide. Protometabolism may have been the original defining feature of emerging life, followed by RNA replication and then modern enzymatic metabolism. Since they combine the physical, geological, and chemical conditions likely to have driven prebiotic synthesis, hydrothermal vents could have provided an ideal environment for the origin of protometabolism. Most of the protometabolic energy of protocells involved the downhill transport of electrons available in reduced compounds to more oxidizing conditions such as molecular oxygen. The general metabolic pathway across all organisms suggests an early origin of protometabolism in the prebiotic world.

Now, I will explore theories regarding the nonenzymatic origin of metabolism and different possible protometabolic pathways in the vent environment. Catalysts may have played a crucial role in polymerization, establishing the early forms of metabolism that ultimately led to the biosynthesis of protein. Modern metabolic pathways most likely emerged through stepwise recruitment of ever more effective catalysts in primordial chemical reaction networks. Metal ions of Fe, Mn, Zn, and Cu were also available in the vent environment, helping mediate catalysis [15–17]. The creation of small organic molecules from inorganic compounds, including mineral-mediated synthesis, was probably the stimulus for metabolism's origin.

7.5.2 The Role of Mineral Surfaces in Protometabolism

The water of the vent environment was rich in ferrous iron and transition-state metals, such as ions of magnesium, copper, and zinc, compounding the catalytic capabilities of the iron-rich clays of the crater. The mineral surfaces of the crater could absorb and concentrate thin films of organic solutes from aqueous solutions. In the 'iron–sulfur world,' Günter Wächterhäuser suggested that metabolic pathways could have begun in vent environments when iron–sulfur (FeS) compounds catalyzed the formation of organic molecules [15]. Little mineral clusters of ferrous iron and sulfide, known as FeS clusters, persist at the core of many enzymes today, suggesting that these minerals could have been catalyzed during protometabolism. In Wächterhäuser's metabolism-first view, when all of life's essential biomolecules were initially manufactured, iron was a dominant dissolved ingredient in the vent environment, which then formed a complex compound with sulfide, and finally precipitated. He specifically proposed that the primordial energy source for the prebiotic fixation of CO_2 involved the oxidation of the mineral pyrrhotite (a form of FeS) to form pyrite (FeS_2). Thus, iron sulfide acts as both a catalyst and a source of energy in prebiotic synthesis:

$$FeS + H_2S \rightarrow FeS_2 + H_2 + \text{energy}$$

The energy released from this reaction causes hydrogen to react immediately with the carbon dioxide in the water to synthesize organic molecules such as formic acid (HCOOH):

$$CO_2 + H_2 \rightarrow HCOOH$$

and

$$CO_2 + FeS + H_2S \rightarrow HCOOH + FeS_2 + H_2O$$

In this simple reaction, a carboxyl (-COOH) group is formed in an extremely simple anabolic reaction by harnessing environmental energy. Carboxyl groups are involved in the linking of amino acids to proteins. Pyrite surfaces would have catalyzed a variety of other reactions.

The crystalline faces of common rock-forming minerals, such as pyrite and montmorillonite, are likely to have played several essential roles in protometabolism [15–17]. They provide a means to concentrate and assemble amino acids to form polypeptides: when a solution containing amino acids evaporates on using a clay substrate; other studies have documented the adsorption and polymerization of amino acids on varied crystalline surfaces [18]. The interface between water and mineral surfaces and free energy available in the vent environment was the main driver for condensation reaction.

The origin of the evolution of metabolism has recently been promoted by the behavior of ZnS, which can harvest sunlight and convert it to form chemical bonds of dicarboxylic acid from CO_2, thus providing the core reactions of universal metabolism before the existence of enzymes [19]. Zhou et al. have suggested that prebiotic metabolites (available from simple sunlight) can catalyze clay minerals (i.e., a zinc clay called sauconite). This work demonstrates the power of clay minerals to replicate and the mechanism by which prebiotic metabolites could catalyze their formation. Since clays have proven to be capable of catalyzing the linkage of activated nucleotides into small RNA-like associations, clay minerals might then have played a key role in concentrating and catalyzing the polymerization of key organic molecules such as RNAs and polypeptides. Indeed, pyrites have been shown to promote some of the condensation reactions involved in polypeptide synthesis [20]. Heating and drying are also other robust ways to drive condensation reactions required to produce polymers from simple organic compounds [1]. Amino acids, short peptides, and cofactors were possibly precursors to more complicated enzymes. Although, compared with later molecular RNA or protein catalysts, they may have limited catalytic abilities in both acceleration and specificity, some small molecules can be remarkably effective catalysts.

7.5.3 Chemical Energy for Protometabolism

To maintain their metabolism, living organisms rely on chains of transfer from an energy source. Since life is an energy-consuming phenomenon using free energy to produce an order from disorder, energy flow through molecular systems would have been integral to life's origin. We believe that the source of energy in prebiotic synthesis came from the vents themselves. Hydrothermal vents play a crucial role in the question of life's origin because they continuously harbor far-from-equilibrium conditions where gases like CO_2, N_2, and H_2 are available. In this environment, the H_2–CO_2 redox reaction provides a constant source of energy. In a hydrothermal vent, geochemically generated H_2 is the primary source of chemical energy by electron transport. Hydrogen is an electron donor, and iron, either in solution or present on a mineral surface, is an electron acceptor. Minerals composed of iron and sulfur can act as oxidants and reductants, thereby providing a source of chemical-free energy for primitive metabolic pathways. Many chemosynthetic reactions in hydrothermal vents, such as hydrogen oxidation, sulfur reduction, iron oxidation/reduction, sulfate oxidation/reduction, and methanogenesis, are quite simple in this way [17]. In the presence of high temperature and high pressure in hydrothermal vents, minerals of the iron–sulfur world tend to dissolve. The dissolved atoms and molecules from the minerals themselves become necessary reactants in the prebiotic soup. Sulfur dissolved from the iron sulfide minerals combines with carbon dioxide and water to form thiols and thioesters—reactive molecules that could initiate protometabolism [16].

De Duve suggested a high-energy thioester-based protometabolism, which follows pathways not dissimilar to modern metabolism, in which thioesters also play a crucial role [10]. Thioester molecules were probably ubiquitous in the primordial soup of the vent environment. Esterification between a thiol (R′-SH) and carboxylic acid (R′-COOH) created these molecules under high-acidic and high-temperature environmental conditions (where R′ refers to an alkyl group):

$$R' - SH\left(\text{thiol}\right) + R' - COOH\left(\text{carboxylic acid}\right) \rightarrow R' - CO - S - R'\left(\text{thioester}\right) + H_2O$$

Thioesters are energy-rich, highly reactive compounds that, due to their ATP-like ability to store chemical energy and release it when thioester bonds are hydrolyzed or phosphorolyzed, can be used as an energy currency themselves. Thioesters are obligatory intermediates in several vital processes in which ATP is either used or regenerated and could

have provided the energy for protometabolism before ATP synthesis evolved. The 'thioester world' represents a hypothetical early stage in the origin of life that could have provided an energetic and catalytic framework of protometabolism. In the thioester world, a core protometabolism gave rise to amino acids and their simple peptides. The catalytic activity of these simple peptides would then become instrumental to the formation of RNA. Eventually, thioesters could have given rise to ATP via acetyl phosphate and phosphorylation of substrates.

$$\text{Acetyl thioester} \rightarrow \text{Acetyl phosphate} \rightarrow \text{ATP}$$

Available ATP (adenosine triphosphate) could then provide energy to drive many protometabolic processes [10]. ATP consists of three components: a nitrogenous base (adenine), the sugar ribose, and the triphosphate. At the center of the ATP, the molecule is a ribose molecule. Attached to one side is a nucleobase, adenine, and a string of three connected phosphate groups. The result is an energy-packed molecule that could have performed many different tasks in the emerging protocells.

ATP hydrolysis yields ADP (adenosine diphosphate), an inorganic phosphate ion (abbreviated as P1, short for (HPO4)–2), and free energy. ATP stores energy in its phosphate bonds, and this energy is released when the bonds are broken. Cells constantly build and break down ATP, creating an ATP/ADP cycle.

$$\text{ATP} + \text{H}_2\text{O} \rightarrow \text{ADP} + \text{P}_\text{i} + \text{free energy}$$

This exergonic reaction releases free energy (ΔG). The reverse reaction, the formation of ATP from ADP and P_i, is an endergonic process and consumes as much as free energy as is released by the breakdown of ATP:

$$\text{ADP} + \text{P}_\text{i} + \text{free energy} \rightarrow \text{ATP} + \text{H}_2\text{O}$$

The ATP/ADP cycle couples exergonic and endergonic reactions. When formed, ATP captures free energy and retains it in the chemical bond between the second and third phosphate. ATP then diffuses throughout the cytoplasm. The stored energy is released by hydrolysis to perform some cellular functions, such as activating metabolic reactions or driving ion transport across the membrane.

7.5.4 Chemiosmotic Energy

The use of ion gradients over membranes for energy conservation, as in chemiosmotic coupling, is universal in living systems, but its origin is obscure. In 1961, the British scientist Peter Mitchell hypothesized that a proton motive force was responsible for driving the synthesis of ATP. The major energy-generating process in modern cells is chemiosmosis, the movement of ions (i.e., protons) across a semipermeable plasma membrane and down the electrochemical gradient that can drive ATP synthesis (electron transfer phosphorylation or chemiosmotic coupling). Protons are driven through a special and complicated enzymatic machine, the proton-translocating ATP synthase. A proton concentration gradient drives the formation of the biochemical 'battery' ATP by the ATP synthase enzyme. With the help of proteins embedded in the membrane, the gradient also passively incites the ions to return. As protons move through ATP synthase, ADP is turned into ATP, the primary currency capturing energy from chemical reactions and transporting it where needed. Because metal ions alone can catalyze it without either proteins or membranes, substrate-level phosphorylation probably came before chemiosmosis [3].

Several investigators have suggested that early cells or protocells could have developed ionic potentials across membranes from which energy could be developed in the form of chemiosmotic potential. Koch and Schmidt [21] proposed that hydrogen could have donated electrons to an internal acceptor so that their charge is separated across the bilayer membrane [1–3]. Iron–sulfur complexes might have initially made this transfer through the oxidation of hydrogen sulfide and ferric sulfide to iron pyrites and two [H^2] on the outside of the cell membrane with the concomitant reduction of CO or CO_2 on the cytoplasm. The resulting proton gradient then provided a proton motive force that could drive the active transport of phosphate and high-energy bond formation.

Goldford et al. [22] introduced a novel systems approach to reconstructing early life, arguing that phosphates might not have been readily available on young Earth four billion years ago. They assert that in primordial biochemical networks, energy currencies might have been simpler than ATP. They additionally point out that thioesters are widespread in modern metabolism, primarily as Coenzyme A (CoA) derivatives (e.g., acetyl-CoA). They suggest alternative iron–sulfur-based bioenergetics (such as ferredoxins, 4Fe4S) in ancient phosphate-independent metabolism that may have given rise to ATP in the acetyl-CoA pathway. In the vent environment, the exergonic reaction of H_2 with CO_2 would give rise to thioesters. Goldford et al. thus build upon the thioester world protometabolism of De Duve [10]. The key is the soluble iron–sulfur-containing protein ferredoxin that mediates electron transfer in a range of metabolic reactions, resulting in the following successive stages:

$$\text{Reduced ferredoxins} \rightarrow \text{Acetyl thioester} \rightarrow \text{Acetyl phosphate} \rightarrow \text{ATP}$$

Chemiosmosis could have developed in protocells with the emergence of a semipermeable plasma membrane (see Chap. 12), deriving energy from differences in sodium and hydrogen ion (proton) concentrations across their membranes.

7.6 Conclusions

There must have been an energy-driven process for life to begin in which complex molecules formed from simple ones inside protocells. There were various forms of energy in hydrothermal vent environments, both global and local, accessible during prebiotic synthesis. Global energy sources include high-energy sources such as solar energy, especially ultraviolet radiation, and impact shock waves, which could have driven the initial synthesis of reactive molecules. Local energy sources could have also created the building blocks of protometabolism. Thermal energy introduced by evaporation would promote monomers' condensation reaction into polymers. Thermal energy derived from the vent itself could have concentrated dilute solutions of the prebiotic soup and mixed molecules by convection current. Additionally, protocells could have harnessed electrochemical energy in the vent environment by redox reactions, such as pyrrhotite oxidation, sulfur reduction, and methane production.

Since enzymes were not available before the synthesis of proteins, minerals containing metals such as iron sulfide probably functioned as catalysts for protocells. The iron–sulfur world of the vent environment provides the possibility for the functioning of a protometabolism: When iron sulfide (FeS) combines with hydrogen sulfide (H_2S), it produces pyrite (FeS_2) plus hydrogen gas (H_2) and a jolt of energy. From this, a thioester-based metabolism could have evolved when thioesters eventually gave rise to ATP through a reaction with pyrophosphate in the hydrothermal vent environment. As energy-rich, highly reactive compounds suitable as an energy currency, thioesters may have played a crucial role in the early stage of protometabolism.

Moreover, protocells could have used other mechanisms to tap environmentally available energy and harness it as ATP. In addition to substrate-level phosphorylation of ADP to ATP through the formation of thioesters, however, chemiosmotic coupling was available only during the formation of the plasma membrane. Some researchers claim that a primitive form of chemiosmotic potential could have been developed in protocells. In this scheme, hydrogen would have been used as a source of electrons donated to acceptors so that their charge is separated across the bilayer membrane. Others have suggested that ferredoxins may have given rise to ATP in the acetyl-CoA pathway. What is clear is that with the development of a semipermeable plasma membrane, protocells developed primordial proton motive forces and chemiosmosis couplings.

References

1. Deamer DW (2011) First life: discovering the connections between stars, cells, and how life began. University of California Press, Berkeley
2. Deamer DW (2007) The origin of cellular life. In: Woodruff TS, Baross JA (eds) The emerging science of astrobiology. Cambridge University Press, Cambridge, pp 187–209
3. Deamer DW, Weber AL (2010) Bioenergetics and life's origins. Cold Spring Harb Perspect Biol 2(2):a004923
4. Chyba C, Sagan C (1992) Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origin of life. Nature 355:125–132
5. McKay CP, Borucki WJ (1997) Organic synthesis in experimental impact shocks. Science 276:390–392
6. Patel HP, Percivalle C, Ritson DJ, Duffy CD, Sutherland JD (2015) Common origins of RNA, protein and lipid precursors in a cyanosulfide protometabolism. Nature Chem 7:301–307
7. Shock E (1990) Geochemical constraints on the origin of organic compounds in hydrothermal systems. Orig Life Evol Biospheres 20:331–367
8. Weber AL (1997) The triose model: glyceraldehyde as a source of energy and monomers for prebiotic condensation. Orig Life 17:107–119
9. Weber AL (1997) Energy from redox disproportionation of sugar-carbon drives biotic and abiotic synthesis. J Mol Evol 44:354–360
10. De Duve C (2005) Singularities: landmarks on the pathways of life. Cambridge University Press, New York
11. Baltscheffsky H (1971) Inorganic pyrophosphate and the origin and evolution of biological energy transformation. In: Buvet R, Ponnamperuma C (eds) Chemical evolution and the origin of life. North Holland Biomedical Press, Amsterdam, pp 392–419
12. Martin WF, Sousa FL, Lane N (2014) Energy at life's origin. Science 344:1092–1093
13. Oparin A (1924) Origin of life (translated from Russian). Dover, New York
14. Haldane JBS (1929) The origin of life. Rationalist Annual 148:3–10
15. Wachterhäuser G (1993) The cradle of chemistry of life: on the origin of natural products in a pyrite-pulled chemoautotrophic origin of life. Pure Appl Chem 65:1343–1348
16. Hazen RM (2005) Genesis: the scientific quest for life. Joseph Henry Press, Washington, D.C
17. Lahav N (1999) Biogenesis: theories of life's origins. Oxford University Press, New York
18. Hazen RN, Sverjensky DA (2010) Mineral surfaces, geochemical complexities, and the origin of life. Cold Spring Harb Perspect Biol 2010(2):a002162
19. Zhou R, Basu K, Hartman H, Matocha CJ, Sears SK, Vali H, Guzman MJ (2017) Catalyzed synthesis of zinc clays by prebiotic central metabolites. Sci Rep. https://doi.org/10.1038/s41598-017-00558-97
20. Huber C, Wachterhäuser G (1998) Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces: implications for origins of life. Science 281:670–672
21. Koch AL, Schmidt TM (1991) The first cellular bioenergetic process: primitive generation of proton motive force. J Mol Evol 33:297–304
22. Goldford JE, Hartman H, Smith TF, Segre D (2017) Remnants of ancient metabolism without phosphate. Cell 168:1126–1134

Chemical Stage: The Analog Information System

8

The Universe is reeking with organic matter. You could say that the Universe is in the business of making life—or that God is an organic chemist.

—Cyril Ponnamperuma, from the interview, 'The Seeds of Life,' 1984.

Life's origin is a sequence of cosmo-geochemical events that led to the assembly of the first cells. The geological stage leading to the prebiotic chemical evolution was based on a series of sequential steps, each determined by its own contingent initial conditions. Impacting asteroids, comets, and interplanetary dust particles likely delivered the organic compounds necessary for the synthesis of life. These compounds gradually accumulated in hydrothermal crater vent environments. However, prebiotic synthesis would require not only their presence but also their concentration. The dilute compounds present in the vent environment could be sufficiently concentrated in three possible ways: first, by the annual evaporation cycles of the hydrothermal crater lakes during the summers; second, by heat and convection currents in the hydrothermal vent; and, finally, by adsorption to mineral surfaces on the crater floor.

The stepwise transition from simplicity to emergent complexity is an intrinsic characteristic of life's origin. Several theories of abiogenesis propose that life emerged spontaneously from the self-assembly of organic reactions that were already occurring chaotically in hydrothermal vent environments. The emergence of life would have required the organization of just the right combination of smaller molecules into much larger macromolecules. Certain key complex molecules somehow managed to emerge from these far-from-equilibrium conditions. One key to understanding life's origin is to recognize the critical role of complex geochemical environments that were in disequilibrium. What kinds of chemistries might have generated these complex biomolecules at the base of the earliest life-forms? During this process of selection, many prebiotic molecules would have been randomly synthesized but, ultimately, having no abiding role, would have been rejected. We will start by analyzing the simpler chemical components of emerging life, those small carbon-based molecules that might have been selected, concentrated, and organized into the essential structures of life. This view of life's origin, as a stepwise transition from geochemistry to biochemistry, morphing smoothly into biology, and evolving on surface environments on early Earth, is relevant to the search for extraterrestrial life on Mars [1].

8.1 Carbon Chemistry and Life

The first vital step in abiogenesis was the synthesis and accumulation of abundant carbon-based molecules. Life as we know it is based on six elements: carbon (C), hydrogen (H), oxygen (O), nitrogen (N), phosphorous (P), and sulfur (S). These six elements, typically abbreviated as CHONPS, are the building blocks of life: their covalent combinations make up most of the biological molecules on Earth.

Chemical structures and chemical reactions involve electrons in the outermost shells of atoms. The elements H, O and S, N and P, and C, respectively, require 1, 2, 3, and 4 electrons to form covalent bonds: the lighter the element, the more stable the bonds. However, atoms such as C, N, O, and S can share more than one electron doublet. Double or even triple bonds can also form covalently. Among these atoms, carbon occupies a unique position: it can complete its outer electron shell by donating or, conversely, accepting four electrons. Carbon serves as the elemental basis for life for a host of reasons. It is the sixth-most abundant element in the universe. It is quite versatile in chemical reactions because it can form one or more covalent bonds with hydrogen, nitrogen, and sulfur. Because of its electron arrangement, its atoms can also link together to form long, resilient polymers as well as bond with up to four other kinds of atoms. Hydrogen, oxygen, nitrogen, phosphorous, and sulfur can all form stable bonds with carbon, and the energy it takes to make or break these bonds is nearly the same. This means that carbon can move among these atoms without expending much energy.

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_8

Carbon's ability to be bonded in a variety of ways increases the diversity of possible compounds and allows the formation of the long macromolecules and polymers on which life is built. Like carbon, nitrogen can form various complex compounds, including amino acids, proteins, nucleic acids, chlorophyll, and the energy carrier adenosine triphosphate (ATP). Sulfur and phosphorous may have been involved in prebiotic energy storage and the transfer of molecules. Hydrogen can form extremely weak bonds with other molecules; this bonding capability makes hydrogen ideal for many biochemical reactions. Hydrogen also combines with oxygen to form water—the universal solvent of life and biosynthesis. For this reason, hydrogen and oxygen are the most abundant elements in living systems, which generally contain more than 60% of water by volume. Despite many elements in the periodic table, life is based on the CHONPS elements simply because they were plentiful when life started. These six elements were assembled into sets of simple molecules that, in turn, were linked to form more complex molecules. Before life could start, key building blocks and molecules containing CHONPS were synthesized in space and were delivered to young Earth and accumulated in hydrothermal crater lakes. These building blocks assembled into higher-order structures [1].

Certain organic macromolecules are essential to all life. All living organisms can be grouped into four different types, i.e., four primary classes of carbon-based molecules. These 'biomolecules' are lipids ($CH_3(CH_2)_{14}COOH$, etc.), sugars ($C_6H_{12}O_6$, etc.), amino acids ($C_3H_7O_2N$, etc.), and nucleotides ($C_6H_{12}O_7N_4P$, etc.). These four classes make up the three essential parts of a living cell: the cell membrane, proteins, and nucleic acids.

Lipids are elongated hydrocarbon molecules, including a wide variety of fatty acids or their derivatives, which are not water-soluble. They play an essential role in cell membranes' formation and functions, including storing energy, signaling, and providing structural components.

Sugars form carbohydrates, Earth's most abundant biomolecules. They exist in cells as single sugar molecules, such as glucose, and complex molecules, such as starch and cellulose. Sugars are an essential energy source and, together with phosphates, form the backbone of nucleic acids. They also provide structural support for molecular communication between cells.

Amino acids are the structural units (monomers) of proteins. Up to 20 amino acids can join in a short polymer chain called a peptide; longer chains are polypeptides or proteins. Proteins catalyze the bulk of chemical reactions that occur in the cell. They form many of a cell's structural elements and help bind cells together into tissues.

Finally, a nucleotide is the basic structural unit of nucleic acids (i.e., RNA and DNA). Each nucleotide is constructed from three smaller molecular parts: a five-carbon sugar (ribose in RNA and deoxyribose in DNA), a nucleobase or base (one of five closely related ring-shaped molecules, such as adenine, guanine, cytosine, uracil, and thymine), and a phosphate group. Nucleotides polymerize to form nucleic acids with sugar-phosphate backbone.

8.2 The Crater Vent Environment: A Simmering Cauldron

Organic compounds could have been washed into freshwater impact crater lakes with active hydrothermal systems where they began to concentrate in sufficient quantities to initiate essential prebiotic chemistry. While in the global ocean, the influx of cosmic ingredients would have been extremely diluted, impact craters encased in a high rim would have hosted several more favorable conditions for the origin of life, all concentrated at a single location: hydrothermal vents in central uplifts as sources of energy; convection currents for mixing and concentrating organic compounds; metal sulfides, clays, impact glasses, and zeolites as secondary hydrothermal minerals; and lake sediments with catalytic mineral surfaces [2–4].

8.2.1 The Primordial Soup in Crater Basins

We assert that life and its information systems coevolved in the hot, dark, and isolated environment of hydrothermal crater basins that served as incubators four billion years ago. Cosmic molecules, both big and small, acquired analog information during synthesis in space and performed specific functions. Primitive Earth favored an analog format to begin prebiotic synthesis from cosmic ingredients [5]. Each cosmic molecule had its analog information system embedded, as the study of Murchison meteorites revealed the self-assembly of lipid vesicles [6, 7].

Each crater basin acted as a giant cauldron to mix cosmic molecules with vent chemicals in three distinct layers. Fatty acids and other hydrocarbons could self-assemble into vesicles floating on the surface of the water like an oil slick. Below this section would lie various soluble ingredients, such as hydrogen cyanide, organic acids, alcohols, sulfates, and ferrous sulfide, all mixed together. This is the 'prebiotic soup' layer, also containing reactive vent molecules, such as H_2S, CH_4, NH_3, and H_2S, as well as metabolic energy sources, such as thioester and ATP [4, 8]. The hydrothermal vent's aqueous current and heat energy would have made the primordial soup thick with cosmic ingredients. Finally, the mineral surface on the floor of the basin would form the bottom layer, with pores and crevices to concentrate and polymerize monomers. Minerals such as pyrite, clay, borate, and zeolite could have catalyzed reactions, which would have then

produced several key components for abiogenesis [9–11]. The entire crater lake formed a simmering boiler: heat from the hydrothermal vents churned the basin water, constantly mixing and combining biomolecules at different layers, causing simple chemicals to grow larger and producing more complex ones by combinatorial chemistry, as a chaotic mix of energy sources and organic compounds was released from the vents. As the organic compounds became more abundant over time, they joined to form complex molecules through self-assembly. Several key components of life, such as amino acids, lipids, sugars, phosphates, and nucleobases, ultimately derived from cosmic ingredients, were segregated and concentrated [12, 13] (Fig. 6.5).

8.2.2 Self-Assembly Processes

Origin-of-life researchers have long anticipated that molecular self-assembly processes may have played a critical role in the appearance of life [6]. Certain organic compounds have been shown to react with each other to form more complex molecules. Lipid molecules, when placed in water, spontaneously self-organize into tiny cell-like spheres, called vesicles. Other molecules such as amino acids and nucleotides could have polymerized into molecules like proteins and nucleic acids, respectively. These, in turn, formed larger self-assembled structures such as the double helix of DNA. Thus, the self-assembly of molecular systems would have been the spontaneous emergence of supramolecular architectures from single components; hydrogen bonding and nonpolar forces would stabilize orderly arrangements of certain molecules to form aggregates with novel properties that emerge only in the resulting structures. Before the advent of protein enzymes and DNA, emerging life-forms must have been produced through spontaneous self-assembly processes.

Self-assembled vesicles must have been crucial components of protocells. Asteroids first delivered cosmic lipids and other biomolecules that then accumulated and concentrated in hydrothermal vents. When concentrated above a certain threshold, fatty acids spontaneously form cell-like vesicles that can grow and divide if continuously supplied with fatty acids. These vesicles could have formed the basis of the protocell by adding lipids (forming the membrane; described in the next chapter) and other organic molecules, which then became primitive cell organelles.

8.2.3 The Analog World

The appeal to the analog framework in prebiotic chemistry is that it is easier to synthesize in hydrothermal crater vent environments under abiotic conditions. The origin of life depicts how an ordered living system can emerge from a chaotic assemblage of the building blocks of life in a hydrothermal crater vent environment. It would have required the organization and selection of just the right combinations of the smaller molecules into larger macromolecules by trial and error in a feedback loop. Analog communication might have played a role as a means of coordination, where information is exchanged between molecules for selection in abiogenesis [5]. Many molecules were discarded from the prebiotic synthesis, and few were selected. We started by analyzing the simpler chemical components of emerging life—these cosmic biomolecules that might have been selected, concentrated, and organized into the essential structures of life.

We identified three major steps in chemical evolution along with the analog information system during prebiotic synthesis in a hydrothermal crater vent environment. These are (1) chiral selection of monomers, (2) conversion of nucleobases to nucleotides, and (3) polymerization of monomers such as amino acids and nucleotides on mineral surfaces.

8.3 Chiral Selection of Monomers

One of the fascinating examples of prebiotic molecular selection was the incorporation of handedness. The strong selectivity of chiral (mirror-symmetrical) molecular species, notably left-handed amino acids and right-handed sugar molecules, is one of life's most distinctive signatures. Louis Pasteur, the famous French chemist and biologist, discovered the chirality in mirror-image crystals of tartaric acid in the mid-nineteenth century. He concluded that some molecules essential for life exist in mirror image forms, much like our ;eft and right hands. Laboratory syntheses of amino acids and sugars yield a racemic mixture of chiral molecules (equal amounts of left- and right-handed enantiomers). Biotic processes, however, are based on patterns of homochirality. DNA, RNA, and their building blocks are all right-handed, whereas amino acids and proteins are all left-handed. Pasteur speculated that selectivity or homochirality of life's molecules might be linked to magnetic fields, but the origin of homochirality has remained one of biology's enduring mysteries.

Monomers such as amino acids and carbohydrates that are essential to life were first recruited from cosmic building blocks to begin prebiotic synthesis in hydrothermal crater basins. Amino acids are polymerized into long protein chains, and simple carbohydrates such as ribose sugars link up with phosphate groups to form RNAs. These molecules came from space in a racemic mixture of chiral molecules, but abiogenesis selected just one homochiral molecule. Life is based on patterns of homochirality, such as left-handed (Levo, L) amino acid (L-amino acid) and right-handed

(Dextro, D) ribose sugar (R-ribose). This asymmetry between L-amino acids and R-ribose is a unique signature of life on Earth, but its origin in prebiotic synthesis remains unanswered.

The evolution of chiral molecules allowed the highly efficient development of biomolecular functions. It introduced a complementarity into synthetic processes that meant that if an evolutionary process was interrupted, it was not lost. Under what conditions, then, did the homochiral characteristics of life originate? This question has intrigued scientists since Pasteur discovered the chirality of biomolecules. No firm conclusion has been reached, but homochirality likely appeared before left-handed amino acids were encoded into proteins and only right-handed sugars formed the backbones of RNA and DNA. Directional selection of the most stable molecules and useful catalytic reactions must have played an essential role in sorting out selected monomers based on their molecular complementarity, specifically left-handed amino acids and right-handed sugars, from highly heterogeneous populations [9, 10]. Significantly, these are the only two chemical species that underwent chiral selection on the pathway to the origin of life. Although mirror versions of these molecules should theoretically work in the same way as their counterparts, their chiral selection may simply have been triggered by a fundamental asymmetry in nature in the early stages of chemical evolution (Fig. 8.1).

Chiral pairs or enantiomers have similar physical properties and act identically in chemical reactions; however, there are stark differences in the interactions between chiral molecules of different types. For instance, the antibiotic penicillin is effective because it only binds to (and destroys) D-alanine, which exists in the cell walls of bacteria but not of humans. Enzymes also distinguish between the two enantiomers of a chiral substrate. Life's strong homochiral preference for L-amino acids and D-sugars at the monomer level must have been fixed in the early history of the abiogenesis of proteins and nucleic acids. Why was the L/D mirror symmetry broken at the molecular level during biogenesis? No one knows, but that distinction would have provided a form of molecular shape recognition during prebiotic synthesis, an analog information system. Homochirality has implications for the optimization of intercellular space during prebiotic synthesis. Racemic or heterochiral life would need the enantiomeric pair of each chiral entity and increase the amount by twofold of the encapsulated molecules. Homochirality also optimizes the stabilization of supramolecular assemblages.

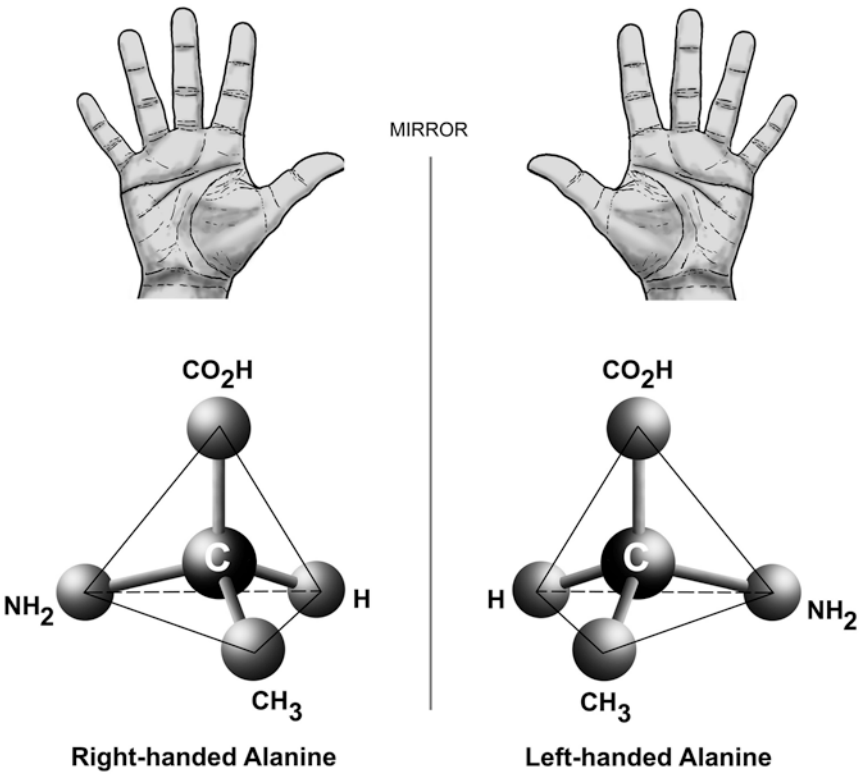


Fig. 8.1 The handedness of life. Our left and right hands are chiral—mirror images of each other. Many organic molecules (such as the amino acid alanine in the figure), called 'chiral' molecules, occur in both left- and right-handed variants, which mirror each other. Chiral molecules arise when four different chemical groups are bonded to a central carbon atom, and these groups can be arranged clockwise or counterclockwise. Chiral molecules have identical chemical compositions but different properties. Living cells contain only left-handed amino acids and right-handed sugars. The selection of homochiral molecules may have taken place in the prebiotic hydrothermal crater vent environment and was probably mediated by certain minerals

Although some amino acids within meteorites were suggested to have an enantiomeric excess of L-amino acids, chiral selectivity most likely would have occurred and amplified in a local chiral environment of a hydrothermal crater vent during prebiotic synthesis. Some researchers argue that chirality occurred as a chance event, resulting from an asymmetric, local physical environment on Earth. Such local chiral environments abounded on prebiotic Earth, both as chiral molecules and in the form of asymmetric mineral surfaces. It is possible that the crystal faces of enantiomorphic pairs of crystals, such as quartz, feldspar, diopside, and calcite, separated chiral molecules from racemic mixtures [9, 10]. Experiments have shown that the chiral faces of these crystals attract chiral molecules of similar handedness. For example, the left-handed faces of calcite may have concentrated left-handed amino acids, and vice versa. The granitic terrain of the crater basins provided various minerals [4]. Perhaps the chiral crystals of the vent environment's mineral substrate facilitated the asymmetry of chiral biomolecules from racemic mixtures of the prebiotic soup. Left- and right-handed mineral surfaces are present on Earth in roughly equal amounts, but the global distribution may not be reflected in the local environment. In the vent environment, left- or right-handed crystal faces might have grafted their chiral bias into the entire biochemical system. Life's chiral bias may have been, quite literally, set in stone.

On the crater floor basin (Fig. 6.5), the mineral surface of clay and pyrite formed—the clay from impact melt and the pyrite from a hydrothermal vent [10]. Another source of clay was the impactor itself, i.e., carbonaceous chondrites. Clays such as smectites drove protometabolism and had catalytic ability. Amino acids and nucleotides are adsorbed on the clay surface on the floor of the crater basin and, subsequently, polymerize [11].

There are several possible ways that a chiral bias may have developed in the vent environment. A soapy film, doped with a chiral oligosaccharide, can separate enantiomers [14]. Perhaps, early lipid membranes, doped naturally with similar chemicals in the vent environment, facilitated monomers' chiral separation. There also exist certain abiotic enzymes, which can resolve racemic amino acids and sugars into their enantiomers.

Why is the L/D symmetry broken by life? Is it a sheer accident in the environmental condition during abiogenesis? Or is chiral purity required for biological function? Is only an L/D protein/carbohydrate combination present for biological function? We do not know the answer. De Duve [15] suggested that the initial choice of R-ribose for the synthesis of nucleotides dictated the choice of L-amino acid for the assembly of peptides. He reduced the choice to 1 ribose molecule instead of 19 amino acids (except for glycine, in which the side chain is H). The molecular choice or attraction might be an example of pure chance or a 'frozen accident.' A chance occurrence of prebiotic chemistry, resulting from an asymmetric, local physical environment, triggered an initial chiral selection.

Experimental evidence corroborates the speculation made by De Duve that R-ribose might have selected the L-amino acid from the racemic mixture that is incorporated into proteins [6]. In protein synthesis, aminoacyl-tRNA synthetase (aaRS) attaches a correct L-amino acid to a specific tRNA molecule to form an aminoacyl-tRNA, which ensures the use of L-amino acid in protein synthesis. Tamura and Schimmel demonstrate that the nonenzymatic aminoacylation reaction of an RNA minihelix has a chiral preference for L-amino acid over D-amino acid [16]. The rationale for using a minihelix in experiments is that it may be a precursor to tRNA and might represent a transitional stage of aminoacylation. Chemical geometry in an RNA minihelix might be the underlying mechanism for chiral selection of amino acid [17].

Asymmetric autocatalysis can drive spontaneous symmetry breaking between L and D enantiomers. The most likely form of autocatalysis in biomolecules is templating of oligonucleotides, as it was shown that homochiral oligomers have a good template, whereas those of mixed chirality do not. This process leads to chiral symmetry breaking when the templated ligation is high [18].

Chirality is an essential component of biochemistry for molecular recognition and replication processes and would seem to be essential for abiogenesis. We speculate that the molecular preference of D-ribose for L-amino acid is linked to one another by stereochemistry and is an early manifestation of analog information [5]. Transmission and amplification of its handedness and its embedded information from the molecular to the supramolecular level are required for abiogenesis. The complementary nature of these two classes of biomolecules was required for creating informational biomolecules: D-ribose for nucleotides and RNA and L-amino acids for peptides and proteins, respectively. Both environmental and chemical factors might have played essential roles in the emergence of homochirality in monomers.

As a side note, although chiral forms of molecules are chemically identical, their chemical actions and biological effects can be surprisingly different. It is extremely crucial for pharmaceutical companies to obtain enantiopure starting materials and separate the racemic mixtures of their final products [19]. Companies routinely select one chiral form over another as more effective than the other, which may be less potent or have dangerous side effects. The left-handed painkiller naproxen, for example, exhibits 28 times the right-handed form's anti-inflammatory activity. The same is true of ibuprofen: the left-handed version is about four times as strong as its right-handed twin. As another example, in the late 1950s and early 1960s, the drug thalidomide was administered to pregnant women to combat morning sickness. Unlike most modern pharmaceuticals, it was sold as a racemic mixture, with disastrous

consequences. The right-handed variant is a mild sedative (thalidomide is still prescribed and is currently approved for the experimental treatment of a variety of diseases, including leprosy, Hansen's disease, Kaposi's sarcoma, myelofibrosis, and human immunodeficiency virus (HIV), but the left-handed variant causes embryonic malformations: thousands of children were born with serious congenital birth defects.

8.4 Conversion of Nucleobases to Nucleotides

Although nucleobases such as adenine, guanine, cytosine, and uracil might have come from space [12, 13] and deposited and concentrated in hydrothermal crater vents, nucleobases such as RNAs use nucleotides for polymerization. Prebiotic synthesis from the assembly of nucleobases, a right-handed ribose and phosphate, is crucial to understanding the origin of life and its information systems.

Becker et al. [20] demonstrated a plausible prebiotic process for the concurrent synthesis of purine (adenine and guanine) and pyrimidine (cytosine and uracil) ribonucleosides, solely driven by wet/dry cycles. In the crater lake environment, a wet/dry cycle was provided by the exposed sloping rim, where each nucleobase was linked to a phosphate group to form a nucleotide. A nucleotide with a sugar-phosphate background became a monomer for polymerization to RNA-like molecules. The wet/dry cycle of the analog information system was instrumental in forming nucleotides that would lead to the origin of RNAs by polymerization (Fig. 8.2).

8.5 Polymerization of Monomers by Condensation Reactions

The next stage of the analog information system is the nonenzymatic polymerization of monomers such as L-amino acids and D-ribose nucleotides into proteins and RNA-like molecules, respectively, by condensation reactions. Peptide and ester bonds link these polymers, respectively. Proteins and polypeptides are both composed of amino acids, but they differ mainly in size: a polypeptide (a string of 2 or more amino acids) is usually much shorter than a protein (at least 50 amino acids in a variety of configurations).

In the prebiotic soup, polymerization is a problematic step in chemical evolution because amino acids and nucleotides do not spontaneously self-assemble into polypeptides and RNAs. We know from the second law of thermodynamics (reviewed in Chap. 6) that energy is needed for polymerization to occur. Complex and highly organized molecules are not expected to spontaneously form from simpler constituents; monomers must absorb energy to link together. If such free energy is not available, then the reaction will not occur. In an energy-rich hydrothermal vent environment, available free energy could have driven polymerization through a condensation reaction (also called 'dehydration reactions' because the newly formed bonds of monomers result in the loss of water molecules). The reverse reaction, called hydrolysis, breaks polymers into monomers by adding water molecules. The water molecules then react with the bonds linking the monomers, separating one monomer after another from the polymer chain.

Both the syntheses of polymers and their hydrolytic breakdown are essential to life. For instance, when we eat food loaded with polymers like starch and proteins, the only way to get nutritional value from that food is to break down the polymers into monomers. Under prebiotic conditions, hydrolysis dominates because it is an exergonic and spontaneous reaction that increases entropy and releases free energy. What conditions could have shifted the reactions in an energetically uphill direction and sustained polymers in the face of a system favoring hydrolysis? In fact, three different processes could have catalyzed polymerization and kept the newly joined polymers from hydrolytic breakdown. Recent experiments have suggested that monomers, incubated by tiny mineral particles such as clay or pyrite, could have polymerized despite these conditions [9–11].

Furthermore, adsorption to mineral surfaces subsequently protects polymers against hydrolysis. Another possible mechanism lies in the wet-and-dry cycles of the fluctuating volcanic pools. Chemical-free energy could have driven a highly concentrated film of monomers to quickly polymerize through a condensation reaction in the dehydrated phase of a wet-and-dry cycle [21]. Other condensation agents in the vent environment remain a third possibility. I discuss these three mechanisms in greater detail below.

8.5.1 Polymerization of Monomers on Mineral Surfaces

Many of life's essential polymers form from water-soluble monomers. The production of such polymers requires two distinct steps: the correct molecules must first be concentrated and then assembled into the desired structure. However, essential biological monomers—amino acids, sugars, phosphates, and nucleotides—are water-soluble and tend to break down, not form, in water. The mineral surface of a crater lake provides an alternative mechanism for polymerization in a hydrothermal setting. In water, monomers such as amino acids and nucleotides can bind to mineral surfaces and undergo polymerization analogous to solid-phase synthesis. Once formed, these various polymers could have been further compounded through chemical reactions into RNAs and polypeptides, both of which

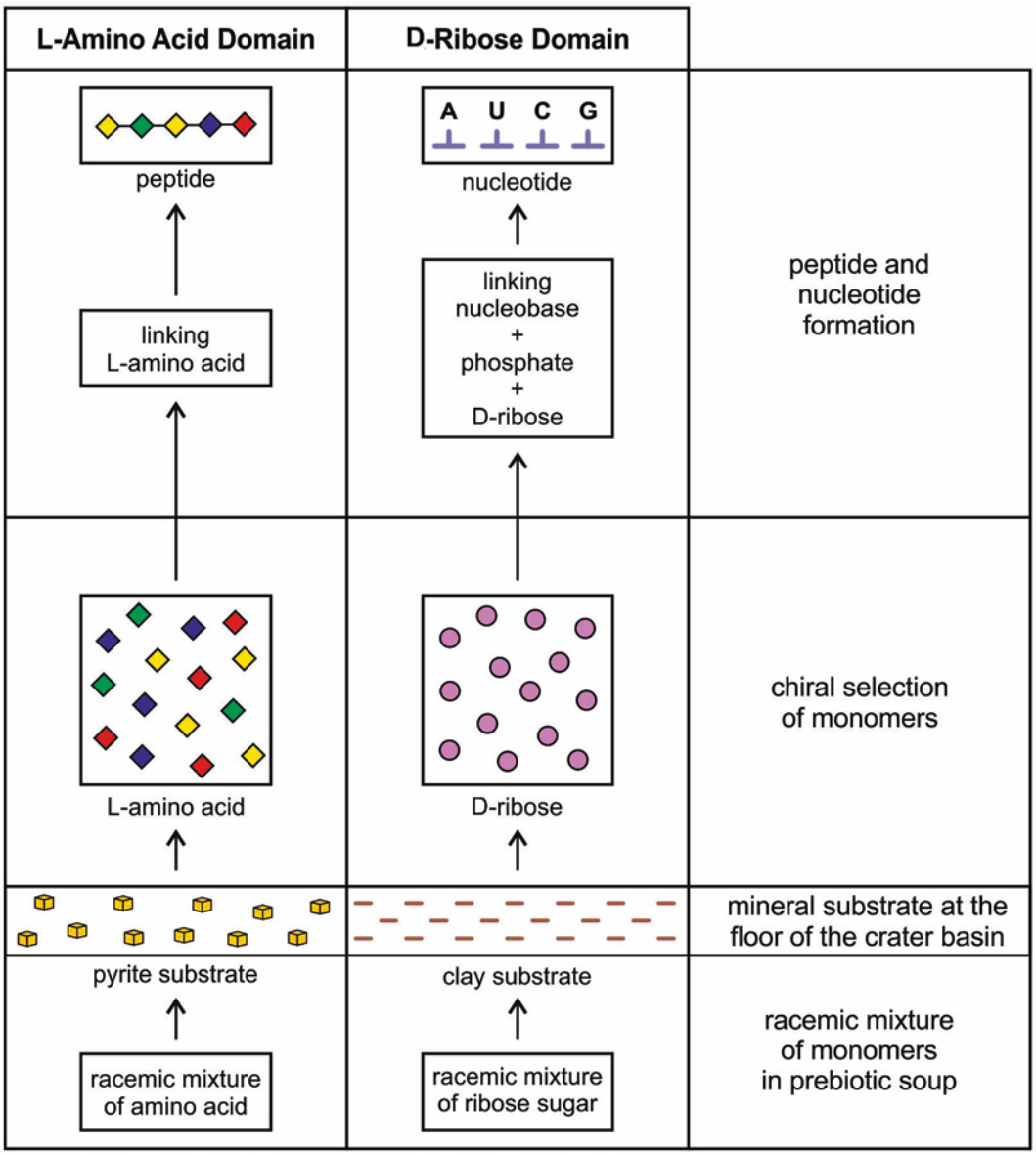


Fig. 8.2 Chiral selection of monomers such as L-amino acid and D-ribose sugar from the racemic mixture on the mineral substrate floor of the crater vent environment. A short chain of the peptide can be formed by linking a few L- to each other via peptide bonds by condensation reactions. L-amino acids become the monomer of proteins. On the other hand, D-ribose joins with a phosphate molecule to form the backbone of a nucleobase; these three molecules join to form a nucleotide, the monomer of RNA

would be necessary for biogenesis. Leslie Orgel, who dubbed this hypothetical process 'life on the rocks,' developed a theory for their formation, noting that when a polymer binds to a mineral surface, the free energy of binding increases linearly as monomers are added to the polymer chain [22].

Molecular adhesive forces between the microscopic layers of minerals first brought disparate monomers close together, allowing them to link and form polymers. Mineral catalysts for polymerization such as clays and pyrites abounded in hydrothermal vent environments, and experiments have shown that condensation reactions, which, as we have noted, were far less likely to occur in the aqueous solution of the vent environment, occur readily when growing polymers attach to small mineral particles [8–11, 23]. These minerals could thus have provided the scaffolding upon which monomers formed polymers such as RNAs and polypeptides.

About 70 years ago, the British biophysicist John Desmond Bernal proposed that clays could have played a key role in initiating condensation reactions since they possess unique surfaces that can adsorb and concentrate

prebiotic organic compounds [24]. This idea of clay surface for polymerization was modified and elaborated by Kloprogge and Hartman [11]. Subsequent experiments have supported Bernal's speculations: when amino acids concentrate and polymerize on clay minerals, they form small protein-like peptide molecules in fluctuating environments [25]. Pyrite surfaces also readily induce peptide formation because the peptide-forming reaction involves a relatively simple mechanism [22]. In experiments modeling underwater hydrothermal vents of early Earth (Ni, Fe), S surfaces converted amino acids into peptides by activating CO. Another recent study has suggested that the clay montmorillonite could have driven the polymerization of nucleotides into polynucleotides [26]. Other researchers have induced clays to act as scaffolds for the formation of RNAs [27]. This body of experimental evidence supports the idea that the mineral surface of a crater lake would have been an important site for chemical evolution: the dilute monomers of the lake were concentrated here and then polymerized.

8.5.2 Polymerization Driven by Evaporation

Fluctuating volcanic pools rich in cosmic monomers were subjected to periodic wet-and-dry cycles resulting from precipitation and evaporation. Polymers could have first accumulated on the surface of the sloping rim of a crater lake to form highly concentrated films. During a dry cycle, when dehydration was occurring, the amphiphiles self-assembled into multilamellar structures capable of concentrating monomers between layers. Under these conditions, water molecules would have evaporated, and ester or peptide bonds were then synthesized. The activation energy for this reaction could have been provided by the elevated temperature of the site. In the hydration (or wet) cycle, the interaction between the water and the dry multilamellar matrix could then generate vesicles [21]. In a similar manner, lipids could also have catalyzed the polymerization of nucleotides into RNAs in these repeated cycles of dehydration and rehydration.

8.5.3 Polymerization by Condensation Agents

Specific nitrogenous molecules (cyanamide, cyanogen, cyanoacetylene, etc.) can act as condensation agents, driving the condensation reactions that link monomers together. We know that hydrogen cyanide and anhydrous liquid ammonia, both of which have been detected in interstellar space, were most likely to have been present in hydrothermal craters. Nitrogenous dehydration reagents could even have induced the formation of polyphosphates and polyphosphoric esters, critical biological condensation agents, and the precursors of ATP.

Polymerization of Amino Acids

There are 20 amino acids in living systems, each with a unique R group (Fig. 8.3a). Prebiotic polypeptides probably consisted of a repeating sequence of L-amino acids available in the vent environment. Their primary structures were certainly simpler than the proteins that sustain life today. Amino acids have a simple core structure consisting of an amino group, a carboxyl group, and a variable R group attached to a carbon atom. They form a polypeptide when a covalent bond forms between the carboxyl (C) group of one amino acid and the amino (N) group of another, which removes water (Fig. 8.3b). The C–N bond that results from this condensation reaction is called a peptide bond. When peptide bonds link a series of amino acids, they are referred to as 'residues' (Fig. 8.3c–e).

Polymerization of Nucleotides

A nucleotide consists of three components: a phosphate group, a D-ribose sugar, and a ring-shaped nitrogenous nucleobase, all of which possibly came from cosmic ingredients (Fig. 8.1e). The phosphate is bonded to the sugar molecule, which, in turn, is bonded to the nucleobase to form a nucleotide [20]. Nucleobases have four nitrogenous bases: adenine (A), guanine (G), cytosine (C), and uracil (U). The sugar-phosphate backbone of the nucleotide forms the structural framework of RNA. The prebiotic synthesis of nucleotides could have occurred in several ways. The simplest way,

Fig. 8.3 (continued) (**b**) The resemblance of an amino acid to a fish helps differentiate its parts. The three amino acids chosen as examples are incredibly similar: each possesses an amino group (the 'head') and a carboxylic acid group (the 'tail'). However, they differ in the 'dorsal fin' (the R group of amino acids), which determines the kind of amino acid (here, alanine, glycine, and serine). (**c**) Three molecules of amino acids can polymerize into a polypeptide by linking the amino group of one to the carboxylic acid group of another. This reaction forms a water molecule through the combination of a hydrogen ion (H^+) discarded from the carboxyl group and a hydroxyl group (OH^-) discarded from the amino group. (**d**) Formation of a longer chain of amino acids (i.e., a polypeptide) by removing a water molecule from each link; a messenger RNA (mRNA)-directed protein molecule is also formed similarly by linking amino acids in the ribosomes during translation. (**e**) Nucleotides can join together into an RNA molecule by linking the sugar (S) and phosphate (P) molecules with the backbone of the ribonucleotide bases (B). (**e**) The linking of nucleotides to form RNAs was accomplished by dehydration. (**f**) The three components of a nucleotide (phosphate group, sugar, and nitrogenous base) in more detail

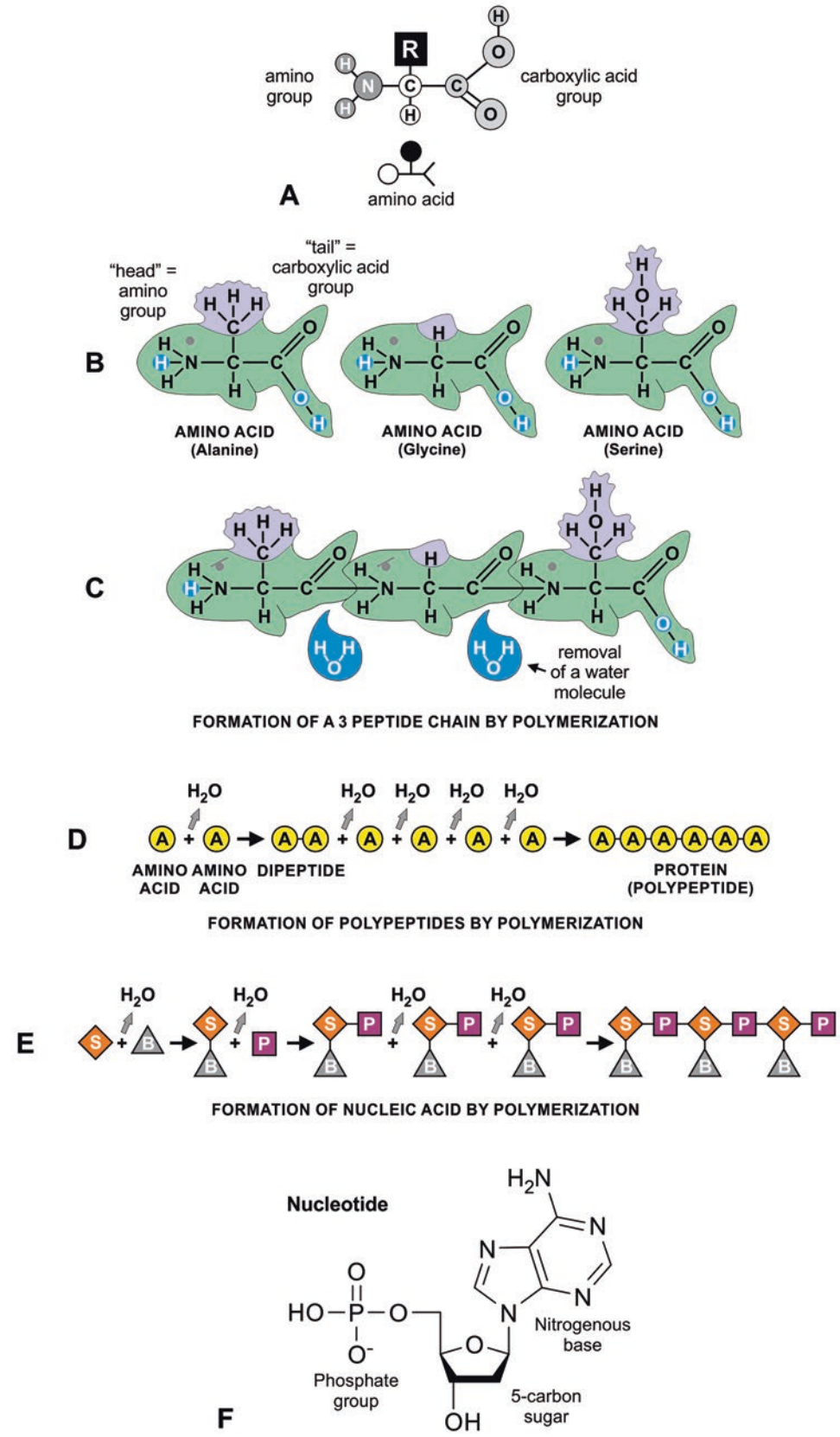


Fig. 8.3 Condensation reaction on mineral surfaces, where activated monomers drive endergonic polymerization reactions. (**a**) An amino acid structure; all amino acids have the same general configuration: a central carbon, bonded to an amino acid functional group, a carboxyl-functional group, a hydrogen atom, and a side chain or R group. (**b**–**d**) Polymerization of amino acids to formpolypeptides by peptide bonds.

conceptually, would be to use a nucleobase, couple it to cosmic ribose, and, finally, phosphorylate the resulting nucleobase.

RNAs can be assembled end to end into linear molecules of nucleotides, just like amino acids when polymerized into polypeptides (Fig. 8.3e). The polymerization reaction involves forming a bond between the phosphate group of one nucleotide and the hydroxyl group of the deoxyribose sugar component of another, resulting in a phosphodiester bond. This bond, like the peptide bond that joins amino acids, results from a condensation reaction and thus removes a water molecule from the bonded nucleotides.

Here, we distinguish two kinds of polymerization of RNA molecules: nonenzymatic (or abiotic) and enzymatic (or biotic). Before the emergence of protein enzymes, all RNA molecules had to be abiotically synthesized. These abiotic RNAs would have been noncoding RNAs, meaning that they lacked the genetic triplet code of biotic RNA and could not encode proteins. They represent 'quasispecies' such as ribozymes out of which many species of RNAs (pre-tRNAs and tRNAs, ribosomes, bridge peptides, pre-aaRSs and aaRSs) may have developed. RNAs within such a pool can bind amino acids available in hydrothermal vent environments. These noncoding RNAs would play critical roles in building the translation machine step by step with the development of the hybrid information system [5].

Polymers such as simple chains of RNAs and polypeptide molecules, driven by convection currents, could have been assembled in significant concentrations in the nanopores of clays and stacking sequences of the mineral substrates of pyrites on the crater floor. The hydrothermal fluid flux, which connected one mineral niche to the next, potentially brought RNAs and proteins into contact with one another. The building blocks of biotic life could then be found in close proximity.

8.6 Conclusions

This chapter offers an account of the chemical evolution of the essential components of life embedded into the analog information system. The six biotic elements (CHONPS) were initially combined into four important molecules: lipids, amino acids, sugars, and nucleotides. These, in turn, were assembled into three major components of cells: lipid membranes, proteins, carbohydrates, and nucleic acids. How did these biomolecules initially come to be, and what drove their synthesis? We believe that hydrothermal crater lakes possessed a combination of hot water and cosmic ingredients, and information sources favorable to the initial concentration of monomers, as well as abundant sources of energy to drive abiogenesis. The convection current arising from the hydrothermal vents could eventually have brought these disparate parts together. Three processes operated at three chemical stages: chiral selection of monomers, conversion of nucleobases to nucleotides, and polymerization of monomers on the mineral surface.

References

1. Sasselov DD, Grotzinger JP, Sutherland JD (2020) The origin of life as a planetary phenomenon. Sci Adv 6:9. https://doi.org/10.1126/sciadv.aaax3419
2. Cockell CS (2006) The origin and emergence of life under impact bombardment. Philos Trans R Soc B361:1845–1856
3. Osinski GR, Tornabene LL, Banerjee NR, Cockell CS, Flemming R, Izawa MRM, McCutcheon J, Parnell J, Preston LJ, Pickersgill AE, Pontefract A, Sapers HM, Southam G (2013) Impact-generated hydrothermal systems on earth and Mars. Icarus 224:347–363
4. Chatterjee S (2018) The hydrothermal impact crater lakes: the crucibles of life's origin. In: Kolb VM (ed) Handbook of astrobiology. CRC Press/Taylor & Francis, Boca Raton, pp 265–229
5. Chatterjee S, Yadav S (2022) The coevolution of biomolecules and prebiotic information systems in the origin of life: a visualization model for assembling the first gene. Life 12:834. https://doi.org/10.3390/life12060834
6. Deamer DW (2011) Self-assembly of organic molecules and the origin of cellular life. Reports for the National Center for science education, 2 (May–August 2003), pp 20–33
7. Deamer DW, Pashley RM (1989) Amphiphilic components of the Murchison carbonaceous chondrite: surface properties and membrane forming formation. Orig Life Evol Biosph 19:21–38
8. Martin WF, Sousa FL, Lane N (2014) Energy at life's origin. Science 344:1092–1093
9. Hazen RN, Sverjensky DA (2010) Mineral surfaces, geochemical complexities, and the origin of life. Cold Spring Harb Perspect Biol 2:a002162
10. Hazen RM (2001) Life's rocky start. Sci Am 284(4):62–71
11. Kloprogge JT, Hartman H (2021) Clays and the origin of life: the experiments. Life 12:259. https://doi.org/10.3390/life120202559
12. Bernstein MP, Sandford SA, Allamonda LJ (1999) Life's far flung raw material. Sci Am 263:42–49
13. Pizzarello S, Cronin JR (2000) Non-racemic amino acids in the Murchison and Murray meteorites. Geochem Cosmochem Acta 64:329–338
14. Kanyanee T, Jakamunee J, Grudpan K, Dasgupta PK (2010) Doped soap membranes selectively permeate a chiral isomer. J Am Chem Soc 132:18045–18047
15. De Duve C (2005) Singularities: landmarks on the pathways of life. Cambridge University Press, New York
16. Tamura K, Schimmel P (2004) Chiral-selective aminoacylation of an RNA minihelix. Science 305:1253. https://doi.org/10.1126/science.1099141
17. Ando T, Takahashi S, Tamura K (2018) Principles of chemical geometry underlying chiral selectivity in RNA minihelix aminoacylation. Nucleic Acids Res 46:1144–1152
18. Tupper AS, Shi K, Higgs PG (2017) The role of templating in the emergence of RNA from prebiotic chemical mixture. Life 7:41. https://doi.org/10.33390/life7040041
19. Nguyen LA, He H, Pham-Huy C (2006) Chiral drugs: an overview. Int J Biomed Sci 2(2):85–100
20. Becker S, Feldman J, Wiedermann S, Okamura H, Schneider C, Iwan K, Crisp A, Rossa M, Amatov C, Carell T (2019) Unified

prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. Science 366L:76–82
21. Damer B, Deamer DW (2015) Coupled phases and combinatorial selection in fluctuating hydrothermal pools: a scenario to guide experimental approaches to the origin of cellular life. Life 5(1):872–887
22. Orgel L (1998) Polymerization on rocks: theoretical consideration. Orig Life Evol Biosph 28:287–234
23. Wachterhäuser G (1993) The cradle of chemistry of life: on the origin of natural products in a pyrite-pulled chemoautotrophic origin of life. Pure Appl Chem 65:1343–1348
24. Bernal JD (1951) The physical basis of life. Routledge and Kegan Paul, London
25. Lahav N, White D, Chang S (1978) Peptide formation in the prebiotic era: Thermal condensation of glycine in fluctuating environments. Science 201:67–69
26. Ferris JP (2005) Mineral catalysis and prebiotic synthesis: montmorillonite-catalyzed formation of RNA. Elements 1:145–149
27. Mathew DC, Luthey-Schulten Z (2010) Influence of montmorillonite on nucleotide oligomerization reactions: a molecular dynamic study. Orig Life Evol Biosph 40:303–317

The Lipid Membrane: Encapsulating Life

9

Carbon has this genius of making a chemically stable two-dimensional, one-atom-thick membrane in a three-dimensional world. And that, I believe, will be very important in future chemistry and technology in general.

—Richard E. Smiley, Nobel Lecture, 1996

The next stage of the analog information system is the spontaneous formation of the amphiphilic bilayer membrane. Every living cell is surrounded by a double layer of phospholipids and proteins called the cell membrane, or the plasma membrane, which separates the interior of the cell from its outside environment. It is selectively permeable, permitting certain molecules to enter and exit the cell. The plasma membrane provides essential protection for the internal workings of the cell. Self-assembly played an essential role in membrane formation. For example, amphiphilic compounds like fatty acids spontaneously assemble into lipid bilayers. A critical question in prebiotic synthesis is how did the pertinent chemicals get organized into larger systems capable of their own self-assembly and polymerization. The evolution of such systems would require functional chemicals to achieve spatial proximity to each other. Encapsulation by a lipid bilayer could have initially brought these chemicals together.

This chapter will trace the likely origin of the lipid membranes in hydrothermal crater vent environments. The isolation of biomolecules was essential for prebiotic synthesis. The protocell may have achieved this necessary step by encapsulating biomolecules in a simple fatty acid membrane, allowing important biochemical reactions to occur between previously disparate biomolecules. Two layers of amphiphilic molecules could have made primitive membranes from a lipid bilayer no more than 5–6-nanometers-thick. Such a membrane could prevent the biomolecules from getting diluted in bodies of water. Self-assembly and encapsulation are the thresholds of the evolution of abiogenesis. With the advent of protein enzymes, a lipid bilayer membrane would evolve into a phospholipid membrane and then into a plasma membrane with an increasingly selective permeability (a topic we will explore in Chap. 14).

9.1 Molecular Self-Assembly of the Primitive Membrane

Self-assembly occurs spontaneously without any regulatory control. It releases energy (an exergonic process) and does not need anything to support it other than the molecular availability of the components. Self-assembly created many of the macromolecules necessary for life from cosmic ingredients or discarded collections of smaller molecules. The first cell membranes were likely to have formed from simple, single-chain lipids present in the prebiotic environment. The subunits interact due to molecular forces such as covalent bonds, hydrogen bonds, Van der Waals forces, electrostatic interactions, and hydrophobic effects. When molecules are concentrated above a certain threshold, molecular forces can drive the self-assembly of fatty acids into membranous compartments bounded by lipid bilayers if the chain lengths are 10 or more carbons long [1].

Over 80 years ago, the Russian biochemist Alexander Oparin suggested that the cell membrane was the first macromolecule to appear during abiogenesis [2]. Oparin proposed that a coacervate, a tiny spherical droplet of assorted lipids, which forms spontaneously, began to enclose concentrated mixtures of macromolecules in the primordial ocean. Since coacervates resemble living cells, Oparin suggested that these were the precursors to the first life. Dyson humorously summed up Oparin's theory saying, 'Life began with little bags of garbage' [3]. These bags—the membranes—were crucial for prebiotic synthesis. David Deamer, who has spent five decades researching the terrestrial origin of life, proposed that the fatty acid membranes themselves set abiogenesis in motion. As molecules were encapsulated, the surrounding membranes transformed the prebiotic world of disordered chemicals into one teeming with the first cells [1]. In the 1980s, Deamer and his colleague Pashley isolated

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_9

fatty acids, in the form of hollow lipid spheres, from the Murchison meteorite. This experiment suggests that the ingredients for making Dyson's 'bags' (or, more properly, vesicles) were likely available on early Earth [4]. Over the years, Deamer has argued that compartmentalization was the most crucial component for the emergence of life and expanded his membrane-first model into a comprehensive theory for how life emerged from the membrane world [1, 4].

The origin of cellular life presumably occurred by the self-assembly of organic compounds to form encapsulated systems capable of catalyzed polymer synthesis [5, 6]. Fatty acids likely played a key role in holding together the building blocks of this system. Amphiphiles, like simple fatty acids, possess a polar or ionic hydrophilic (water-loving) head and a nonpolar or hydrophobic (water-hating) tail. This configuration allows them to spontaneously self-organize into either a monolayer micelle or a bilayer cell membrane in an aqueous environment (Fig. 9.1). Even fatty acids on their own, such as long-chained carboxylic acid extracted from meteorites, form membranous vesicles in water. Similarly, phospholipid membranes are self-assembled into vesicles in water. Both fatty acids and phospholipids make vesicles of similar size; both are thermally stable and are similar in tensile strength; they have similar permeability selectivity trends for small, uncharged solutes [7].

The primitive cell membrane segregated and protected biomolecules from the environment and enhanced the chemical reactions within protocells [1]. Each cell membrane, about 5 nm thick, would have formed a dynamic fluidic structure capable of selectively controlling which molecules entered and exited the protocell. It is important to note that the lipids that form the bilayer (typical to all biological membranes) are not gene products; therefore, the development of genetic material does not need to precede the development of a membrane, although the development of membranes could have facilitated the development of self-replicating genes by providing a protected space in which to evolve and eventually function.

9.2 Formation of the Lipid Membrane

Because the protocell would need to conduct its metabolism without losing energy to the outside hydrothermal environment, systems of encapsulated molecules would have been essential for life to begin [8–10]. In the peptide/RNA world,

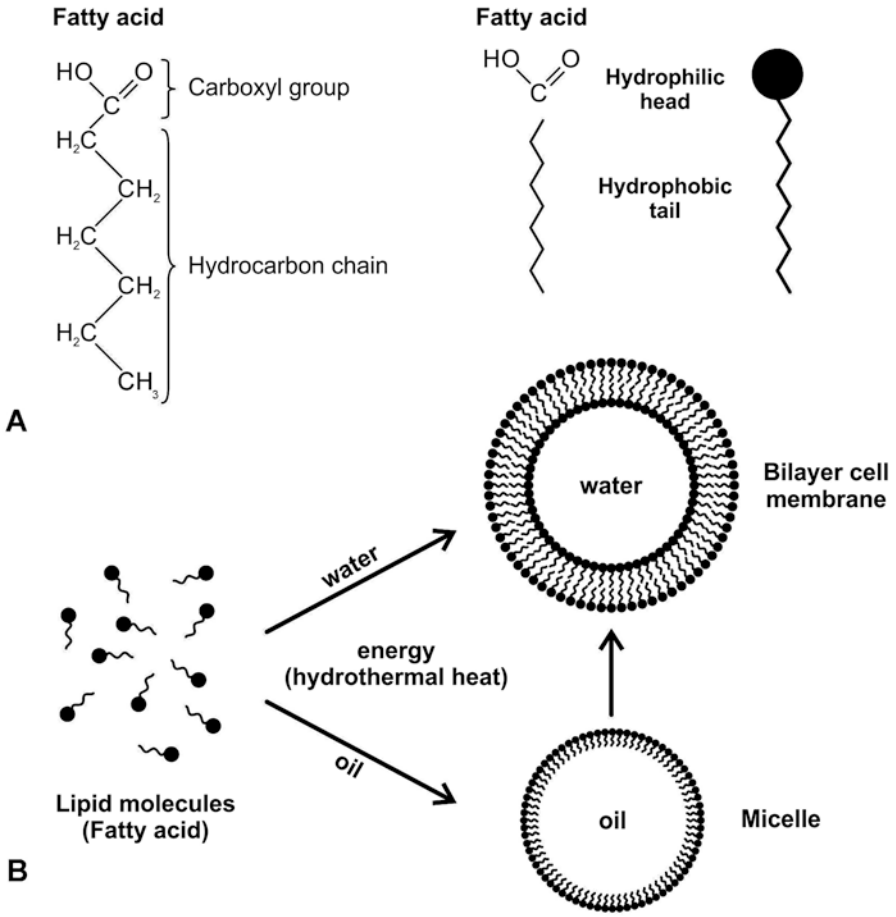


Fig. 9.1 Amphiphilic compounds like fatty acids can self-assemble into cell-sized vesicles bounded by a membrane. (**a**) The polar simple fatty acid was likely a major component of the early prebiotic cell membrane due to its ability to form a vesicle. It has a hydrophilic head and a hydrophobic tail. (**b**) As a monolayer, a micelle can only trap oils, not water, and thus cannot be a precursor to the cell. A bilayer vesicle that trapped water and water-soluble molecules must have given rise to the cell membrane

the primitive cell membrane provided a way to keep peptide and RNA molecules and the translation machinery together at relatively high concentrations [8]. As a prerequisite for maintaining the integrity of cooperative molecular systems that metabolize and synthesize proteins, this compartmentalization created a protected space for internal chemistry differing from the outside environment. Monomers and polymers, now able to remain together, could transfer information between molecules within a protocell. Such membrane-enclosed molecular cooperatives must have preceded the first cells.

9.3 Primitive Amphiphilic Cell Membranes

The secret of membrane construction is the lipid bilayer. Lipid molecules are hydrophobic; they do not mix well with water. They are waterproof and energy-rich. While lipids can spontaneously form a monolayer or a bilayer, due to the polar nature of amphiphiles, only a bilayer could have served as the protocell of the membrane. The hydrophilic head draws the water molecules into the membrane, whereas the hydrophobic tails are attracted to each other because they are repelled by water. The bilayer membranes are stabilized by this hydrophobic effect and the Van der Waals interactions between the tails [11]. A closed monolayer creates a micelle, whose external surface is always composed of hydrophilic heads; the internal surface, consisting of hydrophobic tails, renders a monolayer unable to contain water. A bilayer avoids this by having both exterior and interior surfaces that are hydrophilic. Such a vesicle can trap water and water-soluble molecules such as peptides, ribozymes, RNAs, sugars, and proteins. Life must have arisen from this bilayer of biomolecules.

The simplest example of self-assembling vesicle formation is a soap bubble. Soap is simply a monocarboxylic acid having 10 or more carbon atoms in a hydrocarbon chain that ends in a carboxyl (–COOH) group. When placed in freshwater, lipid molecules aggregate to form huge waterproof sheets, but, when concentrated above a certain threshold, they spontaneously assemble into closed monolayer vesicles. Packing constraints require a minimum length of fatty acids to form vesicles (about 10–20 carbons long) [6, 11, 13]. These newly formed vesicles are flexible, self-sealing, and capable of dividing into two or of joining one another by fusion, all crucial properties for future cell division.

Carbonaceous chondrites—a type of meteorite containing a rich mixture of organic compounds—may have brought to early Earth the fatty acid lipid molecules necessary for the lipid bilayer [1]. Researchers have used organic compounds extracted from the Murchison meteorite to create aqueous vesicles that resemble cell membranes. Unlike actual cells, the vesicles produced were empty but had lipid bilayers composed of heterogeneous mixtures of fatty acids, fatty alcohols, and monoglycerides. Most importantly, the vesicles had enough mechanical strength to have contained and protected critical cell components such as nucleic acids and proteins. Laboratory simulations of interstellar ice mixtures have also produced amphiphilic (that is, having both hydrophilic and hydrophobic parts) vesicle-forming compounds like those found in the Murchison meteorite [4, 6, 11–13]. In simulated syntheses of primitive membranes in the hydrothermal vent environment, Fischer–Tropsch-type reactions produce heterogeneous mixtures of amphiphiles and fatty acids [14]. Thus, the study of carbonaceous materials and laboratory models suggests that vesicle-forming amphiphiles were likely present on early Earth and could have played critical roles in the formation of the bounded membranes required for early protocells.

The long-chain acids and alcohols that contribute to the amphiphilic property of contemporary membrane lipids may be another component of the prebiotic membrane structure. Chain length heterogeneity dramatically lowers the high critical concentration (cac) of fatty acid mixtures, allowing vesicle formation in dilute solutions [15]. These early cell membrane systems would have exhibited many of the characteristics that modern biological membranes possess but without relying on genetically encoded transport systems.

Thus, because they are chemically simple and its common cosmic ingredients and are also prebiotically available, fatty acids are attractive constituents of a protocellular membrane. They can also be further stabilized by the admixture of other simple amphiphiles, such as fatty alcohols and fatty acid glycerol esters. Short-chain amphiphile-based vesicles possess properties similar to those of liposomes formed from phospholipids, which are the primary components of modern cells [6, 7]. Moreover, recent experiments have suggested that while these vesicles are stable in hydrothermal spring water, they cannot assemble in seawater, which lends more support to the conjecture that abiogenesis occurred not in the submarine hydrothermal vent settings but in terrestrial hydrothermal vent environments [16, 17].

9.4 Permeability of the Primitive Lipid Membrane

Modern cells have complete control over their uptake of nutrients and export of wastes through the specialized channel, pump, and pore proteins embedded in their membranes. Membrane-spanning proteins are responsible for the passage of ions, polar molecules, and large molecules that do not readily cross the phospholipid bilayer on their makeup of about 50% of the plasma membrane. In the fluid structure of the plasma membrane, phospholipids and proteins move

back and forth. Membrane-mediating proteins thus provide the selective permeability crucial for metabolism. However, because proteins, phospholipids, and sterols are products of highly complex metabolic pathways that incorporate multiple enzyme-catalyzed steps, they seem unlikely to contribute to the first forms of protocells.

Before the evolution of membrane proteins, intrinsic membrane stability and permeability of polar solutes were the essential features of a primitive cell membrane. As I have suggested, the earliest membranes were most likely composed of simple and readily available fatty acids. Synthetic protocells made of fatty acid membranes reveal that protocells would have retained large polymers while allowing the diffusion of smaller, less polar solutes. The ability to transport polar molecules likely stems from increased lipid dynamics. These laboratory studies show that primitive membranes developed the selective permeability that allowed for the passage of polar solutes, such as nucleotides, which would be necessary for proto-metabolism [18, 19]. How then did this happen?

Passive diffusion could have played an essential role in the nutrient and energy uptake across boundary membranes. In hydrothermal vent environments, both nutrients and chemical energy, including ATP, were available [20]. There is some evidence that suggests that transmembrane diffusion of the lipid cell membrane would have been fast enough to sustain the demands of a primitive metabolism [7]. Fatty acid vesicles are relatively permeable to ionic and polar solutes. Additionally, dynamic fatty acid membranes are permeable to nucleoside mono- and diphosphates and are necessary for spontaneous growth and division [7, 10]. Prebiotic vesicles were undoubtedly composed of complex mixtures of amphiphiles. Compared to membranes composed only of a single molecular species, such as fatty acids, those of mixed amphiphiles often have superior stability and tolerance over a wide range of pH and ionic conditions. Fatty acid/phospholipid blended membranes exhibit high stability to divalent cations (Mg^{2+}) while simultaneously maintaining their permeability to small-charged molecules. This kind of blended membrane system could have been a potential site of protocellular evolution [18].

9.5 Encapsulation of Polymers

The next stage of analog information is the encapsulation of polymers for prebiotic synthesis inside protocells. The encapsulation of RNAs and polypeptides along with amino acids and nucleotides must have occurred very early in the development of life, soon after the availability of these polymers in the hydrothermal vent environment. Encapsulation first requires that the bilayer membrane be open, allowing larger molecules to enter. Polynucleotides and polypeptides cannot permeate lipid membranes. How then were they encapsulated within the lipid bilayer? Two different processes have been suggested for the encapsulation of polymers during terrestrial prebiotic synthesis. In the first model, continuous wet/dry cycles in temporary hydrothermal ponds allow condensation on their surface. The hydration/dehydration cycle permits the formation of vesicles with encapsulated polymers in a hydrothermal pool rich in fatty acids [3, 16, 17]. Upon the completion of the dehydration cycle, amphiphiles could self-assemble into dried multilamellar structures containing monomers between their layers. At the same time, condensation reactions polymerize amino acids and nucleotides into polypeptides and RNAs, respectively, all the while preserving the lipid bilayer. In the following hydration cycle, vesicles are produced when the water interacts with the dry multilamellar matrix. Some vesicles would then contain the polymers, whereas others are empty.

Clay mineral on the crater floor at the mineral/water interface catalyzes the assembly of both the RNA and the membrane vesicle. Clay mineral accelerates the spontaneous conversion of fatty acid micelles into vesicles. In this model, the vesicle crowded on the crater floor on the mineral surface could trap the RNA and polypeptide from the adsorbed surface of the clay and encapsulate them, thus bringing these two components together to generate a protocell-like structure [18, 19].

I had previously suggested an alternative model for the encapsulation of RNAs and polypeptides on the mineral surface of the hydrothermal vent environment (Fig. 8.3) [22, 23]. As we discussed in Chap. 7, mineral surfaces would have been able to concentrate and polymerize monomers and thus produce RNAs and polypeptides. The hollow lipid membranes would stick to the mineral substrate like tiny blisters, providing access to a wide range of polymers and other biomolecules. In this model, the vesicle crowded on the crater floor on the mineral surface could trap the RNA and polypeptide from the adsorbed surface of the clay and encapsulate them, thus bringing these two components together to generate a protocell-like structure [23]. As these protocells were released from the mineral surface, their polymers became encapsulated, ready to participate in further chemical reactions. This initial cooperation between the encapsulated polymers might even be called the origin of life. Within these newly formed protocells, genetic material could reside and be replicated and metabolism could occur. From there, protocells could begin to develop other biotic functions, such as the self-assembly of boundary membranes, transport of monomers, and encapsulation of polymer systems capable of growth and of developing an information system [21] (Fig. 9.2).

Once formed, small molecules and polymers may have crowded protocells. This crowding could have affected

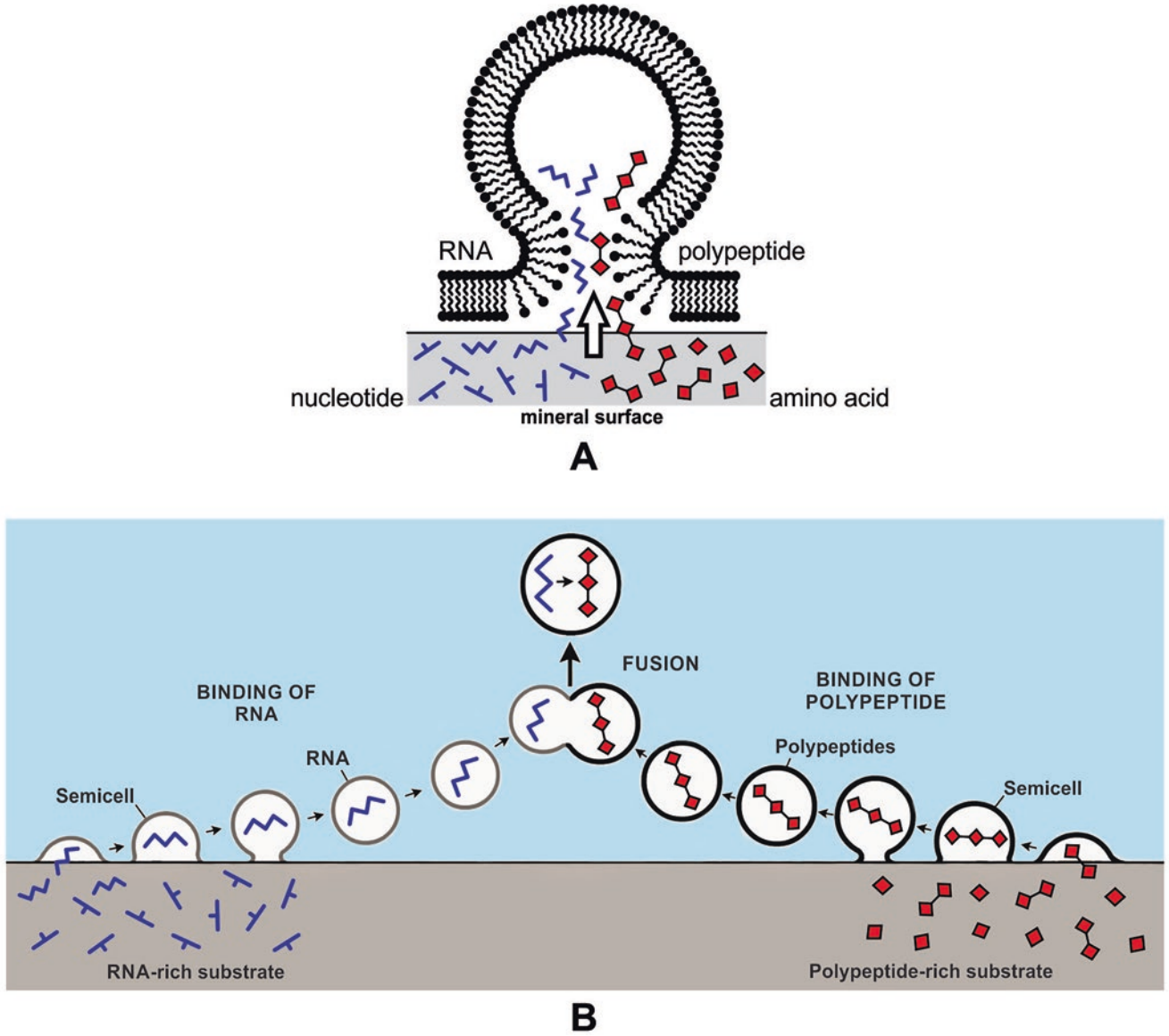


Fig. 9.2 Two possible models of the encapsulation of polymers by simple fatty acid membranes on the mineral surface. In model **A**, both RNAs and polypeptides are brought together in the same vesicle. In model **B**, RNAs and polypeptides are encapsulated separately on the crater basin, then fused in the aqueous environment

reaction rates, changed the structure and activity of water, and even increased protocells' capabilities. It has been suggested that lipid membranes benefit from encapsulating RNA, a growth advantage over empty vesicles [24]. Confinement inside protocells could improve RNA aptamer activity. Vesicle encapsulation could also possibly increase the RNA's fitness landscape in the peptide/RNA world by the chaperon effect. Protocellular organization thereby leads to a direct benefit for the RNA. Moreover, encapsulated molecular crowding might have enhanced protocells' capabilities for evolutionary innovation through the creation of extended networks in the fitness landscape [25]. The membranes of self-assembled compartments have the potential to maintain concentration gradients of ions, thus providing a free source of energy that can drive otherwise unfavorable reactions.

During encapsulation, the vesicles would capture polymers like RNAs and peptides and the prebiotic soup to maintain the crater vent environment. This was the beginning of primitive cytoplasm, an aqueous medium, inside the protocell. This primitive cytoplasm became the built-in reservoir of various polymers and other chemicals when needed (Fig. 9.3). Encapsulated systems of molecules would be essential for abiogenesis in a protected environment, allowing natural experiments for building complex molecules. This was the beginning of the age of protocells. However, such a lipid membrane should have a crucial weakness, in

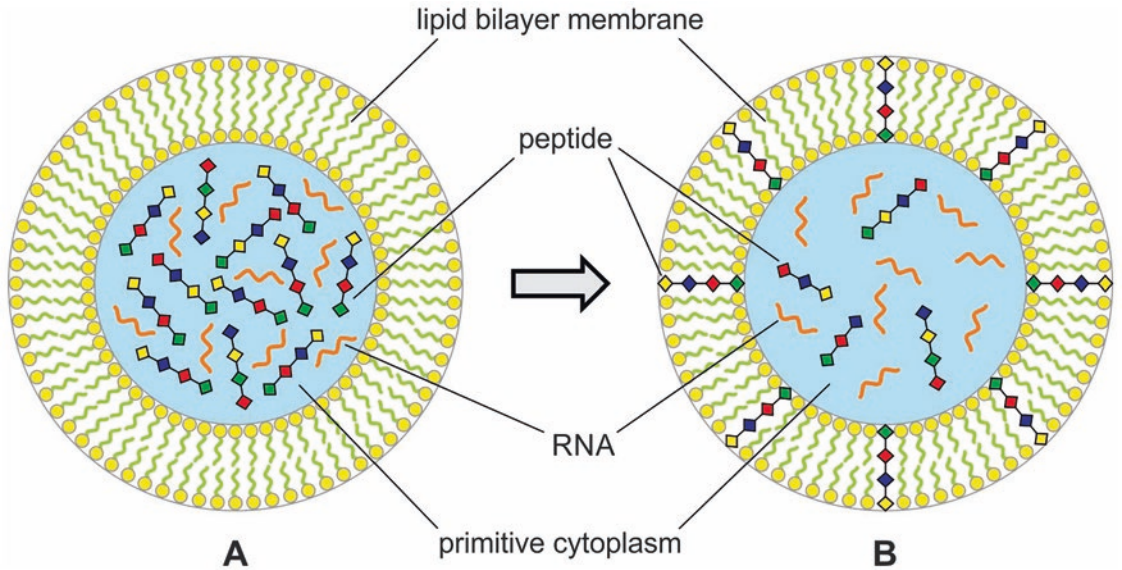


Fig. 9.3 A primitive protocell enclosing assemblages of peptide and RNA molecules. (**a**) Encapsulated polymers such as peptides and RNAs along with the prebiotic soup created primitive protoplasm inside the protocell for prebiotic synthesis in a protected environment. (**b**) Some encapsulated peptide molecules were inserted into the lipid bilayer to enhance permeability in the protocell. The peptides would produce ion-conducting channels through the bilayers that allow a variety of bioenergetics, such as phosphate, thioester, ATP, and nutrients such as amino acids, CH_4, H_2S, CO_2, and NO_2 from the hydrothermal crater vent environment to enter the cell. Molecular crowding inside the cytoplasm would encourage symbiotic relations between peptides and RNAs

that the protocell surrounded by the lipid membrane cannot survive for a long time because it would be difficult to incorporate enough hydrophilic compounds through the lipid membrane because of its poor permeability. The deficiency of the lipid membrane was improved with the development of peptide channels that allowed hydrophilic compounds and other nourishment from the environment to enter the protocells. If the protocells incorporated RNA molecules, then they could undergo a primitive form of growth and division [28].

9.6 Insertion of Peptides into the Lipid Bilayer Membrane

Lipid bilayers were a barrier to diffusion for water-soluble solutes such as amino acids and phosphates or simple ions like sodium (Na^{+1}) and potassium (K^{+1}) [26]. The bilayer barrier is essential to maintain the integrity of polymers, but nutrients from the environment by diffusion were also necessary for the growth of protocells. In an elegant experiment, Hladky and Hayden [27, 29] suggested a mechanism by which the permeability of protocells can be enhanced by inserting peptides such as antibiotics called gramicidin. These pore-forming peptides can spontaneously insert across a lipid bilayer. Some peptides could enter the lipid layers in prebiotic synthesis to create a channel for ions and soluble solutes (Fig. 9.3). Deamer [11] elaborated this concept of peptide insertion to enhance permeability in protocells. This insertion of peptides into lipid bilayers could have been a precursor to the modern cell's plasma membranes, in which proteins are inserted into the phospholipid bilayer to transport a given solute across the bilayer barrier. At this stage, a rudimentary form of signal transduction had evolved from the environment to the cytoplasm via the peptide channel.

9.7 Growth and Division of Protocells

Membrane-bound protocells containing a set of monomers and polymers could grow and divide. Such protocells could acquire resources and energy from the environment. Laboratory simulations hint at a solution to the primitive cell division mechanism. Lipid vesicles extracted from the Murchison meteorite underwent spontaneous primitive cell division in the laboratory, with no external forces acting upon them [28]. When a mixture of these cosmic vesicles, amino acids, and nucleic acids was shaken, the vesicles trapped the organic molecules inside them and began to interact. This intake suggests that vesicles can take substances from outside themselves through their lipid walls to build new walls and new contents. A large vesicle mimics a primitive kind of cell division. With the development of the peptide channel, the protocells could grow and divide.

The physics of 'chemically active' droplets, which cycles chemicals in and out of the surrounding fluid, may shed light on the origin of protocell division [29]. The team studied a

theoretical model for the behavior of a liquid droplet in a chemically disequilibrated system. This 'active droplet' behavior differs from passive and more familiar tendencies of oil droplets, which join to form bigger droplets without dividing. On the other hand, these chemically active droplets can grow to a stable size by taking resources from the environment. Droplet growth eventually leads to instabilities—linked to the changing shapes of the droplets. The droplet keeps elongating and pinches in at the middle, which has low surface areas. Eventually, surface tension causes it to be split into a pair of droplets (Fig. 15.5). This process of dividing droplets somewhat mimics the spontaneous vesicle division in the Murchison meteorite [28].

In a laboratory simulation, a genome-rich vesicle increased in size at the expense of an empty vesicle. When its greater size imposed too much osmotic stress, pearling instability developed, and the stretched vesicle divided into two, with each daughter vesicle retaining some of the original genomic contents [30]. A recent work on model protocell membranes has demonstrated that vesicles can grow as filamentous structures and divide spontaneously under mild shear forces. With photochemical stimulation, a robust 'pearling' mechanism produces many small daughter vesicles [25]. Self-replicating membranes can divide spontaneously or under the influence of external environmental forces [4], and high environmental shear forces can cause vesicles to divide. In a similar way, a protocell with cytoplasm can divide into two daughter cells but the cytoplasm division in the daughter cells may be unequal.

Synthetic biologists use simple 'protocells' to study the origin of cell division, but previous models could not reproduce both the genome and the membrane sustainably. Kurihara et al. [31] proposed a recursive self-proliferating model protocell that represents a step toward the eventual production of model protocells that can mimic cell division. They used a novel system, fusing the self-reproducing vesicles with feeder vesicles, thus allowing the vesicle composition to be sustained over multiple generations. Because of competition, the larger vesicle grows more quickly and fuses with the feeder vesicles. Therefore, feeding the protocells by vesicle fusion offers a practical pathway for indefinite self-reproduction (Fig. 9.4).

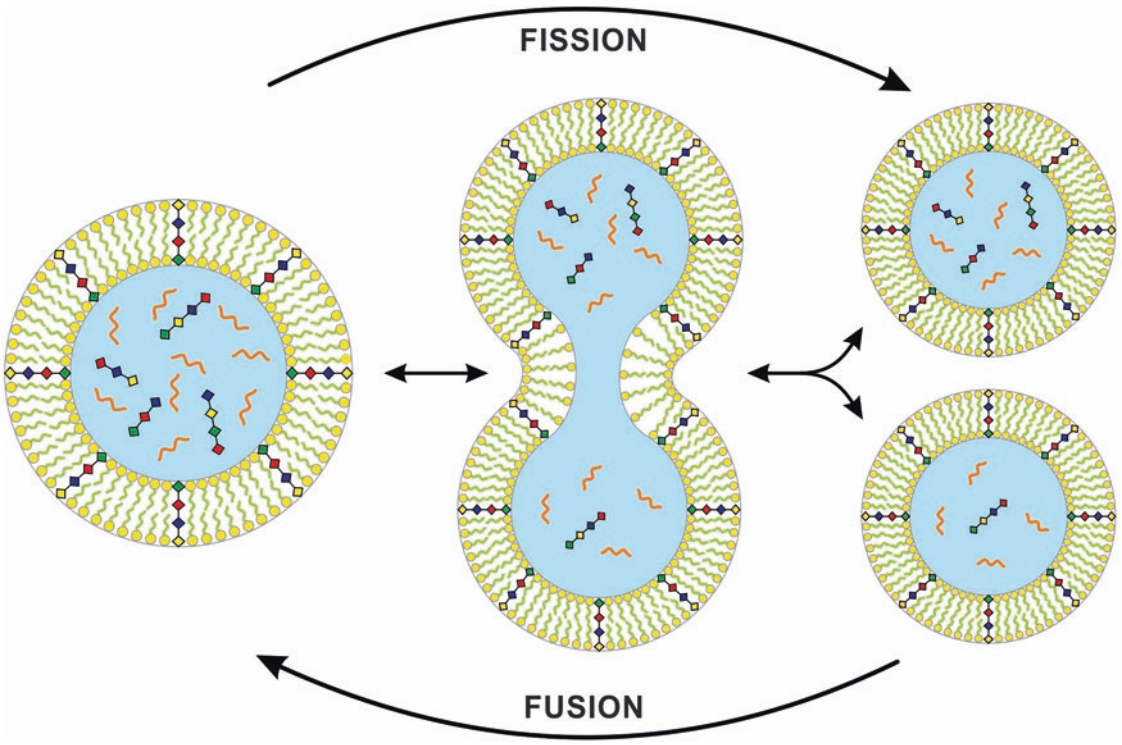


Fig. 9.4 Fusion and fission of lipid bilayers with an inserted peptide channel (see Fig. 9.3 for an explanation). The inserted peptide channel allowed nutrients and bioenergetics from the hydrothermal vent environment to enter protocells by diffusion for growth and division. These protocells form flexible, semipermeable vesicles, capable of dividing into two such daughter vesicles or of joining with others, without losing their structural continuity at any moment. Unlike living cells, the division of protocells is asymmetric, where daughter cells might inherit an unequal amount of cytoplasmic content. The transfer of information from parent to daughter cells is vertical. The cellular division of the first cells inherited this property of the lipid bilayer vesicle

9.8 The Phospholipid Membrane

There are two stages of membrane evolution: the primitive lipid membrane evolved in the peptide/RNA world and the more sophisticated phospholipid and plasma membranes emerged after the advent of RNA-directed proteins. We will discuss the evolution of phospholipid and plasma membranes in the section of 'The Advent of Proteins' (Chap. 14). Protocells used both self-assembly and directed assembly processes to grow and evolve. Self-assembly was essential to the synthesis and stability of membrane structures and protein folding. Directed assembly underlies the synthesis of proteins according to the base sequences of messenger RNA (mRNA). At the beginning of the information stage, the translation system and the genetic code evolved step by step in the peptide/RNA world in the enclosed cell membranes [5]. Once the mRNA-directed protein synthesis became established, various enzymes were created to meet the catalysis demand and metabolism. The next step involved the encapsulation of a more complex enzyme system capable of catalyzing fatty acids to form phospholipids. The transition from single-chain lipids to phospholipids was triggered by the availability of a wide range of enzymes in the protein/RNA world that catalyzed the conversion of lipid membranes into phospholipid membranes.

9.9 Conclusions

During the Late Heavy Bombardment period, carbonaceous asteroids delivered a variety of organic compounds that would become the basis for the lipid bilayer membrane. In freshwater conditions such as hydrothermal crater lakes, a mixture of amphiphiles self-assembled into stable membranous compartments bounded by lipid bilayers. Their permeability could have been improved if peptides were inserted into bilayers, allowing passage of nutrients and small ions for metabolism and growth. These peptides would have produced channels capable of conducting ions through bilayers, allowing phosphates and other nutrients to enter the cell. The insertion of a peptide into a bilayer may have been the precursor to the plasma membrane in which proteins form transmembrane channels that conduct ions. These vesicles could have encapsulated the necessary polymers (too large to cross the assembled membrane) in two ways, through the wet/dry cycle at the water/air interface or by the capture of polymers in semicells at the mineral/water interfaces, where mineral surfaces formed a breeding ground for both polymers and bilayer vesicles. Bilayer membranes would keep molecular systems together for prebiotic synthesis. Protocells with peptide channels could grow and divide, thus mimicking the first cells, but unlike living cells, the division of protocells is asymmetric with its genome content.

References

1. Deamer DW (2011) Self-assembly of organic molecules and the origin of cellular life. Rep Natl Center Sci Educ 2(May–August 2003):20–33
2. Oparin AI (1924) Origin of life (translated from Russian). Dover, New York
3. Dyson F (2004) Origins of life. Cambridge University Press, Cambridge
4. Deamer DW, Pashley RM (1989) Amphiphilic components of the Murchison carbonaceous chondrite: surface properties and membrane forming formation. Orig Life Evol Biosph 19:21–38
5. Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. Nature 409:387–390
6. Deamer DW (2017) The role of lipid membrane in life's origin. Life 7(1):5. https://doi.org/10.3390/life37010005
7. Mansy SS, Schrum IP, Krishnamurthy M, Tobe S, Treco DA, Szostak JW (2008) Template-directed synthesis of a genetic polymer in a model protocell. Nature 454:122–126
8. Chatterjee S, Yadav S (2019) The origin of prebiotic information system in the peptide/RNA world: a simulation model of the evolution of translation and the genetic code. Life 9:25. https://doi.org/10.3390/life9010025
9. Ross DS, Deamer D (2016) Dry/wet cycling and the thermodynamics and kinetics of prebiotic polymer synthesis. Life 6:28. https://doi.org/10.3390/life6030028
10. Mansy SS (2009) Model protocells from single-chain lipids. Int J Mol Biol 10:835–843
11. Deamer DW (2019) Assembling life: how can life begin of Earth and other habitable planets? Oxford University Press, New York
12. Deamer DW (1997) The first living systems: a bioenergetic perspective. Microbiol Mol Biol Rev 61:230–261
13. Dworkin J, Deamer D, Sandford S, Allamandola I (2001) Self-assembling amphiphilic molecules: simulated in simulated interstellar/precometary ices. Proc Natl Acad Sci 98:825–819
14. McCollom TM, Ritter G, Simonetta BR (2003) Lipid synthesis under hydrothermal conditions by Fischer-Tropsch-type reactions. Orig Life Evol Biosph 29:153–166
15. Monnard PA, Deamer DW (2002) Membrane self-assembly processes: steps toward the first cellular life. Anat Rec 268:197–207
16. Mishteyn D, Damer B, Havid J, Deamer D (2018) Amphiphilic compounds assemble into membranous vesicles in hydrothermal hot spring water but not in seawater. Life 8(11). https://doi.org/10.3390/life8020011
17. Damer B, Deamer DW (2015) Coupled phases and combinatorial selection in fluctuating hydrothermal pools: a scenario to guide experimental approaches to the origin of cellular life. Life 5:872–887
18. Jin L, Kamat NP, Jena S, Szostak JW (2018) Fatty acid/phospholipid blended membranes: a potential intermediate state in protocellular evolution. Small. https://doi.org/10.1002/wmll.201704077
19. Budin I, Prwyes N, Zhang N, Szostak JW (2014) Chain-length heterogeneity allows for assembly of fatty acid vesicles in dilute solutions. Biophys J 107. https://doi.org/10.1016/j.pbj.2014.07.067
20. Lane N, Martin WF (2012) The origin of membrane bioenergetics. Cell 151:1406–1416
21. Hanczyc MM, Fujikawa SM, Szostak JW (2003) Science experimental models of primitive cellular components: encapsulation, growth, and division. Science 302:618–622
22. Chatterjee S (2015) The RNA/protein world and the endoprebiotic origin of life. In: Clarke B (ed) Earth, life, and system. Fordham University Press, New York, pp 39–79
23. Chatterjee S (2016) A symbiotic view of the origin of life at hydrothermal impact crater-lakes. Phys Chem Chem Phys 18:20033–20046

24. Saha R, Pohorille A, Chen IA (2014) Molecular crowding and early evolution. Orig Life Evol Biosph 44:319–324
25. Saha R, Verbanic S, Chen IA (2018) Lipid vesicles chaperone an encapsulated RNA aptamer. Nat Commun 9:2313. https://doi.org/10.1038/s41467-018-04738-8
26. Mansy SS (2009) Model protocells from single-chain lipids. Int J Mol Biol 10:35–843
27. Hladky SB, Hayden DA (1972) Ion transfer across lipid membrane in the presence of gramicidin, A: studies of the unit conductance channel. Biochem Biophys Acta 274:294–312
28. Deamer DW (2011) First life: discovering the connections between stars, cells, and how life began. University of California Press, Berkeley
29. Zwicker D, Seyboldt R, Weber CA, Hyman AA, Jülicher F (2017) Growth and division of active droplets provides a model for protocells. Nat Phys 13:408–413
30. Chen IA (2009) Cell division: breaking up is easy to do. Curr Biol 19:R327–R328
31. Kurihara K, Okura Y, Matsuo M, Toyota T, Suzuki K, Sugawara T (2015) A recursive vesicle-based model protocell with a primitive model cycle. Nat Commun. https://doi.org/10.1038/ncomms9352

The RNA World: Reality or Dogma?

10

It is clear that at some early stage in the evolution of life, the direct association of amino acids with polynucleotides, which was later to evolve into the genetic code, must have begun.

—Leslie Orgel, 1968

In living organisms, three kinds of polymers enclosed by plasma membranes—DNA, RNA, and proteins—underpin nearly everything that happens inside cells. The semipermeable plasma membrane protects the cell from the outside environment and regulates materials entering and exiting the cell. DNA encodes the digital information that RNA translates into proteins, and proteins organized into cell structures do everything else, from building and repairing cell components to driving metabolic reactions.

The progenitor to all life was a self-replication entity capable of evolving. However, what were the very first self-replicators that were directly ancestral to all life? The RNA world is a hypothetical stage in the origin of life, in which the self-replicating molecule RNA proliferated before the emergence of lipid membranes, peptides, proteins, and DNA. Like DNA, RNA stores and transmits information. Like proteins, RNA ribozymes can catalyze specific biochemical reactions. In the RNA world scenario, the earliest forms of life may have relied solely on RNA to catalyze chemical reactions and store genetic information. Despite extensive research, however, the link between nonenzymatic RNA polymerization and RNA self-replication remains unsolved. In this chapter, before we discuss the popular RNA world hypothesis, we will briefly outline the molecular structure of RNA molecules and our theory of their origins in the hydrothermal vent environment.

10.1 Structure of the RNA Molecule

Ribonucleic acid (RNA) is a linear molecule with a ribose–phosphate backbone and four different nitrogenous bases: adenine (A), cytosine (C), guanine (G), and uracil (A). Cytosine and uracil are classified as pyrimidines, whereas adenine and guanine are purines. Purines always bond with pyrimidines via hydrogen bonds such that the four RNA nucleotide bases form the canonical Watson–Crick base pairs: A always pairs with U and G always pairs with C. Each nucleotide base consists of a D-ribose sugar, a phosphate group, and a nitrogenous base (Fig. 9.1b). Although generally single-stranded (Fig. 10.1c), RNA can form localized double-stranded regions in hairpin loops by bending back on itself. RNA is both a messenger for the genetic information carried to the translation machinery during DNA replication and, as a component of ribosomes, a catalyst for protein synthesis.

The ribose sugar of RNA is a pentose ring consisting of five carbon atoms, numbered 1′–5′, and one oxygen atom. A base (A, C, G, or U) is attached to the 1′-carbon end and to a phosphate at the 5′-carbon end. The hydroxyl (-OH) group on the 2′-carbon of ribose sugar (absent in DNA) makes RNA much more reactive and prone to hydrolysis (Fig. 9.1b). This chemical liability of RNA is believed to be one of the reasons why chemical evolution preferred DNA as the carrier of genetic information.

10.1.1 Nonenzymatic Synthesis of RNA

In modern organisms, numerous enzymes participate in producing nucleotides and joining them together, including RNA polymerase that catalyzes and synthesizes RNA in transcription. Since no such enzymes would have existed in the hypothetical RNA world, primeval RNA would have to have been synthesized abiotically from cosmic ingredients. Researchers have detected a wide range of nucleobases, as well as ribose sugar, phosphate, and amino acids, in carbonaceous meteorites [1, 2]. The extraterrestrial ribose sugars may have contributed to forming biopolymers such as RNAs and polypeptides. Additionally, the lack of deoxyribose sugar (the backbone of DNA) in meteorites supports the hypothesis that life began in an RNA world.

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_10

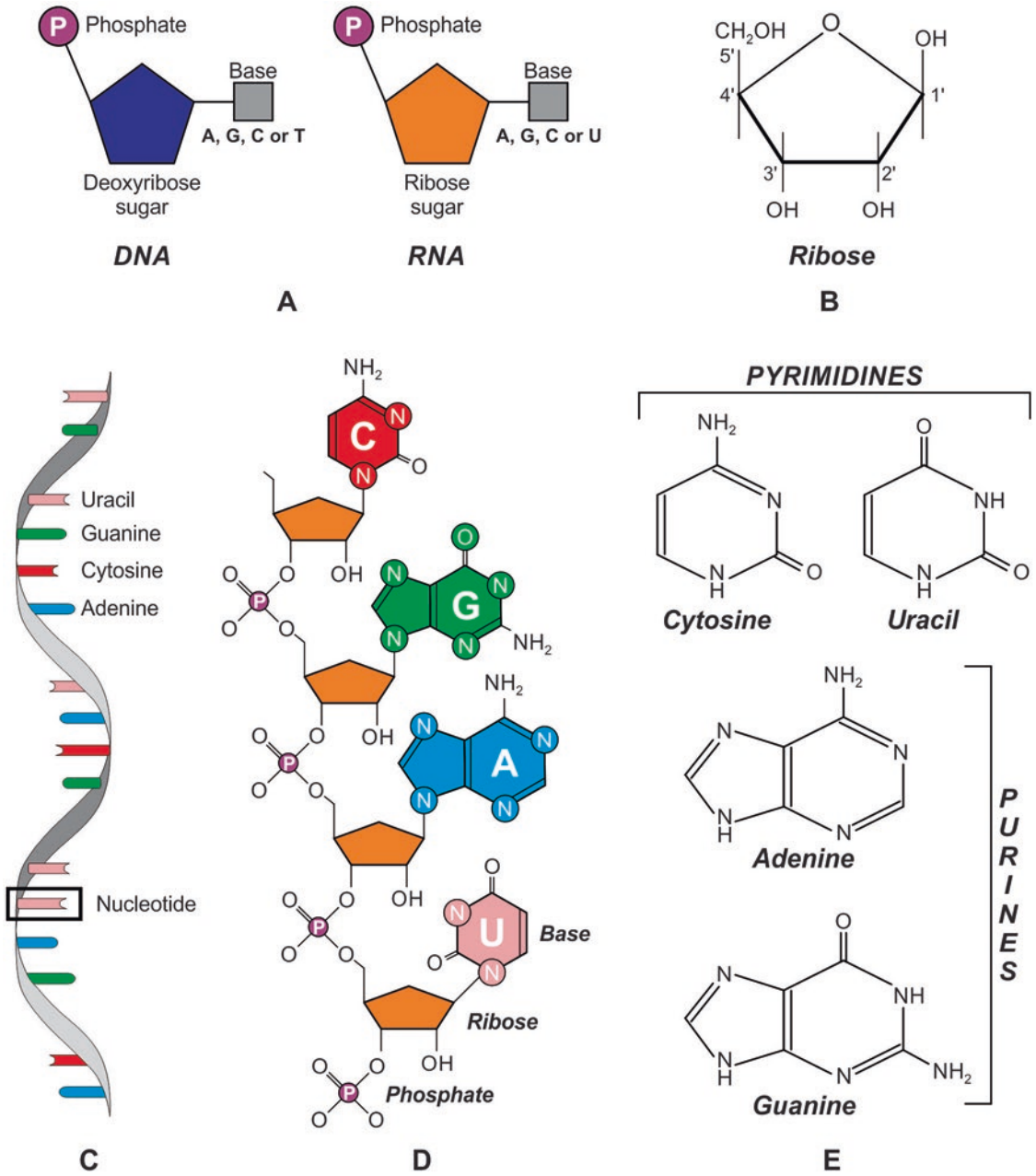


Fig. 10.1 The structural difference between RNA and DNA molecules. (**a**) DNA contains the deoxyribose sugar and thymine base, whereas RNA contains the ribose sugar and uracil base. Uracil differs from thymine, in that it lacks a methyl group on its ring. (**b**) Ribose and deoxyribose are pentose sugars containing five carbon atoms. These are numbered 1′–5′. (**c**) RNA is a single-stranded molecule, whereas DNA is a double-stranded molecule. (**d**) RNA has a sugar–phosphate backbone, where the bases are attached to the ribose sugar. (**e**) among four bases, cytosine and uracil are smaller and have a single ring. and are grouped with the pyrimidines; while adenine and guanine are larger and have two rings and are grouped with the purines

The joining of nucleotides into an RNA chain is one of the most significant challenges in the prebiotic synthesis scenario. Three different catalysts (montmorillonite clay surface [3, 4], lipids [5], and peptides [6]) could have initiated the nonenzymatic synthesis of RNA molecules. All were available in the hydrothermal vent environment. Peptides available on the mineral surface may have aided in the initial polymerization of nucleotides. When researchers activated added nucleotides to a kind of volcanic clay that would have been available in the vent environment, condensation reactions formed RNA molecules up to 55 nucleotides long [3]. If all four nucleotides formed naturally on the mineral substrate, then they could zip together quickly to polymerize an RNA molecule with a backbone of alternating sugar and phosphate groups. The bases attached to the sugar would have constituted a four-letter alphabet in which biological information could begin to take form (Fig. 10.1).

The first RNA molecules were probably random associations of nucleotides in the vent environment. They may have other bases (such as imaginary F and N in the figure) besides adenine, guanine, cytosine, and uracil (Fig. 10.2). However, the chemical complementarity between A and U, on the one hand, and between C and G, on the other, seems to have singled out the ACGU alphabets for information transfer. Such selectivity could provide an answer to why there are four bases poised for the origin of life.

10.2 Replication of RNA

The simplest replicating system consists of one type of molecule that makes the exact copies of itself. In the right chemical environment, such an isolated molecule acts as a template and will multiply itself. Single self-replicating molecules are intrinsically complex molecules. RNA is such an autocatalytic molecule that it can replicate itself. However, in relation to the prebiotic stage of RNA, the term 'self-replication' is oversimplified because RNA carries out two functions—informatics and synthetic. RNA is an informatic molecule because it contains a sequence of bases analogous to the letters in a word. Since the original template of RNA formed randomly in a prebiotic environment, it would not have had much information because the letters would have been random, without any meaning; the only information it would have carried is the base pairing rule. Base pairing depends on structural complementarities such that the two bases fit into each other like puzzle pieces. They are, in fact, flat and adhere due to weak hydrogen bonds. Base pairing determines the shapes of RNAs in the living world and governs all the basic biological information transfers involving those molecules, whether in replication or translation. The synthesis component of RNA replication requires complex building blocks, an energy source such as adenosine triphosphate (ATP), and strong chemical support.

Self-replicating RNA molecules are believed to be crucial to the origin of life as the basis of heritability, a necessary characteristic of living systems. However, the critical copying of a genetic molecule appears to be an exceedingly complex process involving an array of proteins and other cellular components. Was an RNA ribozyme sufficient to catalyze the reactions needed for self-replication on early Earth? The RNA world theory hinges on this vital question.

10.2.1 Base Pairing and Self-Replication of RNA Molecules

Unlike polypeptides, polynucleotides can directly guide the formation of the exact copies of their sequences. This ability depends on the complementary base pairing of nucleotide subunits, allowing one polynucleotide to act as the template for another. The linear sequence of nucleotides in an RNA molecule usually occurs in the form of a single strand made up of a sequence of the four bases. Hydrogen bonds hold the base pairs together, although the G–C pair, with its three hydrogen bonds, is stronger than the A–U pair, which has only two.

Complementary base pairing, also known as 'hybridization,' allows one RNA molecule to act as the template for another to form (Fig. 10.3a). Soon after replication, two double strands of RNA separate into four single-stranded molecules, one of which is identical to the original strand. A single strand of RNA can specify a complementary polynucleotide sequence, a 'flipped' version of the original, whereas the second round of copying restores the original sequence.

Such a complementary mechanism producing more diverse populations of molecules lies at the heart of RNA replication. However, to promote polymerization, such an arrangement requires additional enzyme catalysts that were not available in the RNA world. For years, researchers have questioned whether there might have been a more straightforward way to copy RNA—perhaps by RNA itself.

10.2.2 RNA Replication by Ribozymes

Ribozymes (**ribo**nucleic acid en**zymes**) are RNA molecules that can catalyze specific biochemical reactions in a way like those of enzymes. However, the primary structure of RNA molecules is much more restricted than that of proteins by having only 4 bases versus the 20 types of amino acids at the base of proteins. By bending back on itself, RNA can form localized double-stranded regions in hairpin loops, resulting in a secondary structure (Fig. 10.3b). It can also fold into various complex tertiary structures, i.e., three-dimensional structures that give ribozymes their catalytic ability (Fig. 10.3c). Crystal structures of several ribozymes have provided detailed insights into the folds of RNA molecules. Many ribozymes have either a hairpin- or hammerhead-shaped active center and a unique secondary structure that allows them to cleave RNA molecules at specific sequences. All ribozymes catalyze the cleavage of RNA chains or the formation of bonds between RNA strands.

The RNA world hypothesis suggests that modern metabolism evolved from a system in which RNA served both the genomic role, now filled by DNA, and the catalytic role, performed by proteins. RNA began as a self-replicator: to make a copy of itself, the RNA molecule had to provide a sequence that could be copied. It also had to catalyze polymerization reactions that would link monomers into a copy of the template. This model is corroborated by the discovery that a ribozyme can catalyze a diverse range of chemical reactions. The RNA world hypothesis relies on the premise that in the

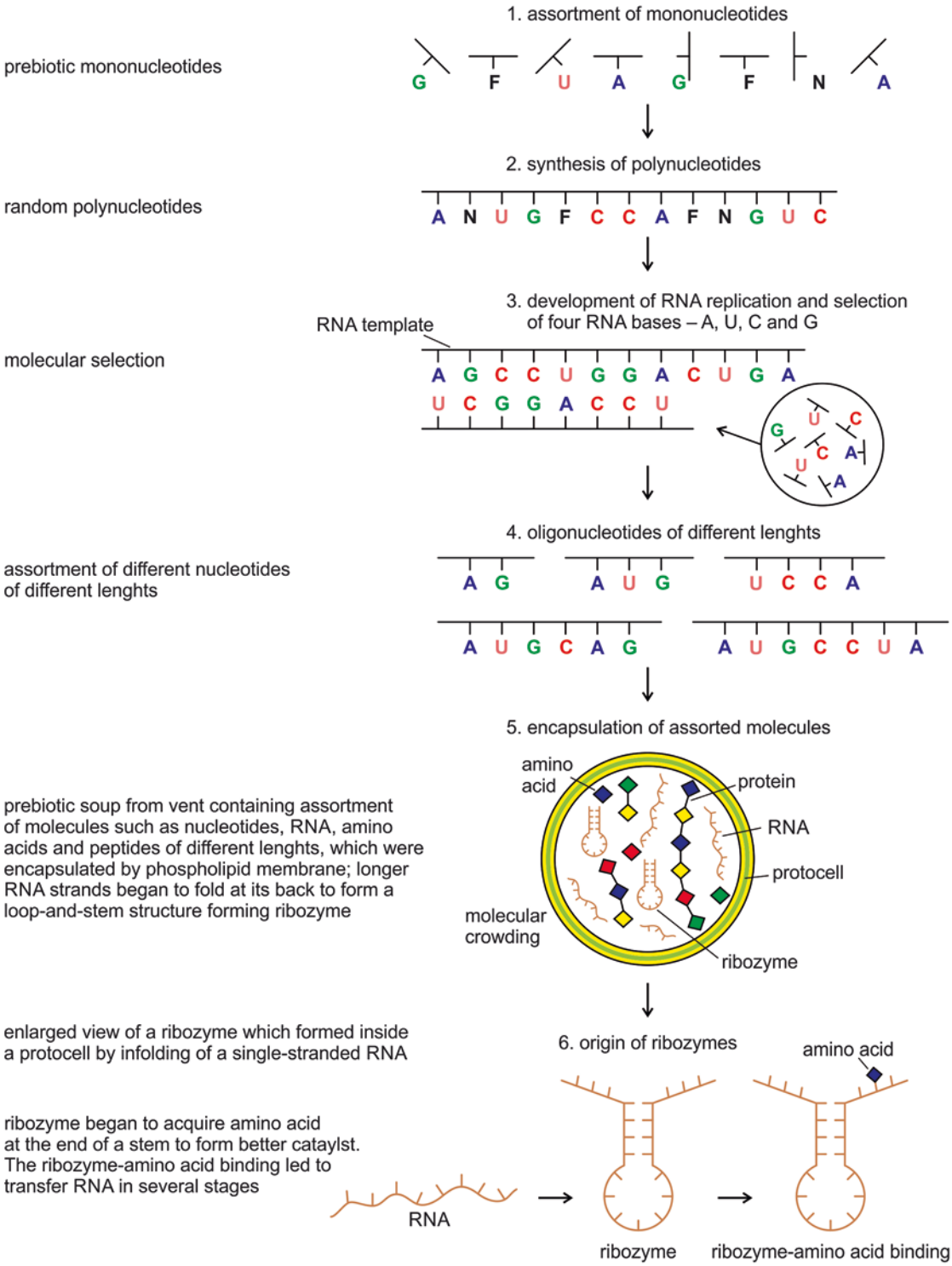


Fig. 10.2 Six main steps represent the early evolution of noncoding RNA in the hydrothermal crater vent environment. There was an assortment of different nucleotides (including some not found in RNA). In the second stage, these nucleotides are randomly assembled into polynucleotides by polymerization through condensation reactions removing water molecules. In the third stage, the selection of four nucleotides—A, U, G, and C—occurred during their replication by Watson–Crick base pairing. In the fourth stage, the nucleotides underwent polymerization, creating a random mixture of polynucleotides of different sequences and lengths. In the fifth stage, an assortment of biomolecules, such as amino acids, mononucleotides, oligonucleotides, and peptides, were randomly encapsulated and crowded in protocells. In response to this crowding, the single-stranded RNAs began to fold, forming the double-stranded stem and single-stranded loop of the characteristic RNA hairpin. In the sixth stage, this secondary RNA structure formed a ribozyme and began to act as an enzyme. Stems were created by hydrogen bonding between complementary base pairs. The ribozyme acquired amino acids, at the CCA sequence of the stem, as 'cofactors,' increasing its catalytic capability. The opposite end of the loop contains three unpaired bases facing outward, which formed a binding site for the attachment of three corresponding mononucleotides. This was the beginning of the emergence of the proto-tRNA

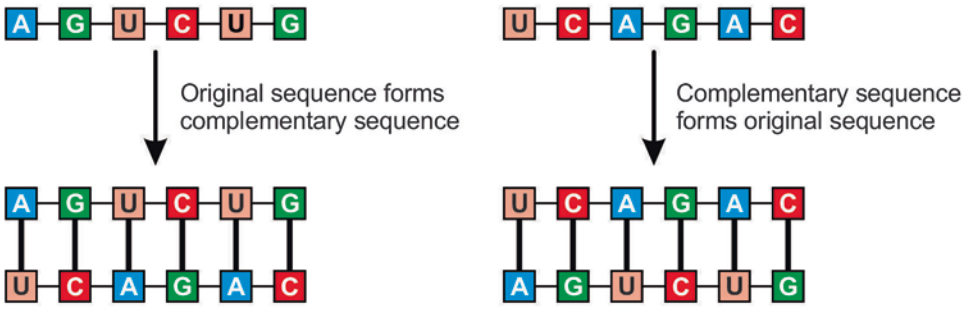


Fig. 10.3 Replication of an RNA molecule by base pairing. In step 1, the original RNA molecule acts as a template to form an RNA molecule of a complementary sequence. In step 2, this complementary RNA molecule itself acts as a template, forming an RNA molecule of the original sequence. Since each templating molecule can produce many copies of the complementary strand, these reactions can multiplicate the original sequence. In the RNA world, this mechanism of replication could not occur in the absence of protein enzymes

early evolution of life, some elusive primeval RNA sequences could perform the function of an RNA replicase, a protein enzyme—they could catalyze RNA replication. However, the absence of an RNA replicase from the RNA world is arguably the most significant limitation of the RNA world hypothesis. Perhaps, in the prebiotic environment, the RNA polymerase ribozyme could have provided some catalytic help to overcome this deficit.

10.3 The RNA World

In the 1960s, not long after the determination of DNA's double helix structure and a couple of decades before the discovery of ribozymes, Carl Woese [9], Francis Crick [10], and Leslie Orgel [11] each independently identified RNA as the primordial molecule. Walter Gilbert coined the phrase 'RNA world' in 1986 [12]. The 'RNA world' represents a hypothetical stage in the development of life before either DNA or proteins existed, with their functions being carried out solely by RNA molecules. The first genes were not DNA molecules but short strands of RNA, which then began self-replicating (an ability that neither DNA nor proteins possess; DNA cannot replicate without prior formation of an RNA primer). The DNA-based life we see today is believed to have emerged from a less complex, now extinct, form of life based on RNA. For decades now, the RNA world hypothesis has been the dominant hypothesis for the origin of life. In the early 1980s, Thomas Cech at the University of Colorado and Sidney Altman at Yale University discovered ribozymes. Cech isolated a ribozyme from the protist *Tetrahymena thermophila*, and Altman found a similar ribozyme in the bacterium *Escherichia coli* [13, 14]. Their discoveries earned Cech and Altman the 1989 Nobel Prize for chemistry. They showed that ribozymes could catalyze chemical reactions, including one that initiated self-replication (Fig. 9.3c). This self-replicating ability of this form of RNA, combined with its ability to store information, quickly led to the suggestion that the first living cells arose in an RNA world.

RNA's dual functions put the RNA world hypothesis on the map, and the discovery of ribozymes was a further milestone in the origin-of-life research. The ubiquity and versatility of RNA's twin abilities to serve both genetic and catalytic roles made it the leading candidate for life's first biopolymer. Eventually, the hypothesis continued, these RNA chains became longer and longer, replicating themselves, catalyzing complex reactions, and beginning to encode proteins, thus ushering in the 'RNA–protein' world. Only later did DNA take over the role of a gene carrier from RNA, and proteins become the major catalysts and structural components of cells [8]. More recently, the crystal structure of a ribosome has revealed that it is, in fact, a ribozyme with a central core of RNA. A ribosome's active site is the ribozyme that catalyzes the peptidyl transferase reaction of protein synthesis, further strengthening the RNA world hypothesis [7].

10.3.1 Artificial Ribozymes

The RNA world scenario is built upon certain assumptions regarding RNA's catalytic capability. RNA polymerase ribozymes would have to have been responsible for replicating ribozymes, including themselves (via their complementary sequences). However, RNA replication involves a complex set of reactions—far more complex than those currently known to be catalyzed by RNA. Since the discovery of ribozymes, researchers have looked to newly synthesized artificial ribozymes for a greater understanding of the possible dynamics of the RNA world.

On the assumption that the primordial RNA replicase was lost during the evolutionary process, the functional aspects of life's first genetic system, as hypothesized in the RNA world scenario, must be studied using modern-day analogs obtained through directed evolution. The best of these

analogs is a family of ribozymes based on the R18 RNA polymerase ribozyme [15]. R18 is a ribozyme, approximately 200 nucleotides long, derived from a ligase ribozyme generated de novo from a random RNA sequence pool. In the search for an elusive RNA ribozyme that exhibits polymerase-like activity, artificial ribozymes have been developed that recognize a primer–template complex.

The techniques used to discover artificial ribozymes center on their dual nature as both a catalyst and an information polymer. Investigators produce vast populations of RNA catalysts using a polymerase enzyme with occasional mutations. Extensive libraries of different RNA strands are examined to catalyze a specific chemical reaction or bind to a particular substrate. The strands that pass the test are separated and reverse-transcribed, selected for their catalytic activity, and amplified. Sequences then subsequently evolve by repeating the process through many rounds. Such techniques develop synthetic ribozymes capable of connecting multiple RNA molecules to generate longer strands or using the nucleotide triphosphate to assemble a new RNA strand from an RNA template [15–19].

An RNA polymerase ribozyme can replicate short RNA sequences in the complete absence of proteins. The new ribozyme can carry out a remarkably complicated and challenging reaction. Laboratory experiments have demonstrated that an RNA-based genetic system would require ribozymes that catalyze their own replication. In some experiments, the ribozyme could grab an RNA molecule as a template for adding nucleotides by adding molecules one at a time onto an RNA fragment. The ribozyme could add on 14 nucleotides in total, with an accuracy of 97%. The best ribozyme created so far—able to replicate an impressive 95-nucleotide stretch of RNA—is 190 nucleotides in length, far too long a sequence to have arisen through any conceivable process of random assembly. The most extended primer extension performed by an RNA polymerase is 220 bases [19]. This new polymerase demonstrates the feasibility of evolving artificial RNA replicase ribozymes in the future. However, none of the ribozymes generated in the laboratory yet possess the sophisticated properties needed to replicate RNA accurately. Despite these in vitro experiments, there is still a lack of convincing evidence to suggest that RNA could catalyze its own replication in prebiotic conditions. The self-copying ribozymes work only if provided with the right oligonucleotide components. What is more, sustained cycles of replication and proliferation require special conditions to ensure that the copies can be separated from their RNA templates. Recently, an RNA in vitro evolution has shown promise that it can copy its own template with low fidelity [20]. Despite the technical achievement of these artificial ribozymes, the odds of an undirected, naturally occurring ribozyme replicase emerging in the real prebiotic world are so low that it is likely that this phantom replicase never existed.

10.4 Flaws in the RNA World

Today, the RNA world hypothesis has been popularized to such an extent that its speculative foundation tends to be forgotten. Developed and elaborated upon by highly respected scientists and in the most advanced genetics laboratories, it is perhaps the most extensively studied model for the emergence of life. Despite promising indirect evidence for the RNA world hypothesis, several significant challenges remain. How the RNA world came to be in the first place remains a mystery [21–30]. The replication of genetic molecules is exceedingly complex, involving many proteins and other cellular components. The existence of ribozymes does not prove the existence of the RNA world but merely suggests that RNA serves both informatics and replicative functions in the prebiotic world. It is still not clear whether ribozymes could have catalyzed all the chemical reactions necessary for life or to what extent they would have been involved in proto-metabolism before protein enzymes took charge of replication [23].

A major deficit of the RNA world hypothesis is a plausible mechanism for synthesizing RNA polymers through non-enzymatic processes. Even under optimal laboratory conditions, it is challenging to make RNA without enzymes. In the chaotic prebiotic environment, nucleotides would have required enzymes to form the first RNA strands. Furthermore, for RNA to function and replicate, primitive RNA would have required polypeptide-like short chains of amino acids or polypeptides themselves to provide some catalytic assistance. Despite considerable experimental and theoretical efforts, the RNA world hypothesis still cannot explain how an efficient RNA replicase or a translation system could have emerged in the prebiotic environment [29–31].

These theoretical inadequacies have proved so intractable that many researchers are increasingly considering a 'pre-RNA world.' Perhaps RNA was not the first replicator after all. Many have begun exploring the possibility that RNA was preceded by a simpler nucleic acid, perhaps peptide nucleic acid (PNA), threose nucleic acid (TNA), or glycol nucleic acid (GNA). These molecules are like RNA, but their basic units are believed to have likely formed spontaneously. They could then have been replaced by RNA. For example, it has been suggested that the earliest life may have used PNA—an artificial nucleic acid in which various pyrimidine and purine bases are linked by methylene carbonyl bonds to the backbone—as a genetic material because it is incredibly robust, simply formed, and can spontaneously polymerize at 100 °C [31]. However, evidence is lacking for the existence of this sort of precursory genetic system in modern biology. The RNA world hypothesis is currently so shaky that in 2009, Thomas Cech, one of the architects of the theory and the discoverer of the ribozyme, disavowed his initial position in favor of an alternative theory, the RNA–protein world [32].

Without a doubt, RNA is an excellent informational molecule, as demonstrated by messenger RNAs and RNA viruses, but the RNA world hypothesis cannot explain the origin of the genetic code, in which base sequences define the amino acid sequences of specific proteins. Associated macromolecules like RNAs and polypeptides are complementary and highly likely to have evolved in tandem. From the beginning, there must have been a division of labor between nucleic acids, the carriers of information, and proteins, the catalysts of replication. It is from the distinct but complementary roles of RNAs and polypeptides that the process of translation and the genetic code came about [28]. The ability of an RNA molecule to act as a catalyst depends on its three-dimensional structure. RNA sequences are highly suitable for storing genetic information, but their simple linear chemistry is not highly suited to catalysis. In contrast, amino acids are more numerous and chemically more diverse than nucleotides and produce proteins capable of assuming an infinite variety of three-dimensional spatial configurations to act effectively as enzymes.

Lacking other molecular companions, the RNA world is a lonely place. RNA molecules are inherently fragile and are easily broken down into their constituent nucleotides by hydrolysis. As we have noted, a central challenge to the RNA world hypothesis is the discovery of RNA replicase, an enzyme that catalyzes RNA using it as a template, a protein molecule that must have been the forerunner of the RNA world. However, there are no naturally occurring ribozymes that can catalyze the necessary chemical reactions [34]. No one has yet found intact RNA in fossils, ancient bones, or pollen grains. Moreover, to perform its many tasks, RNA must be encapsulated [33]. This implies, once again, that membranes arose much earlier than RNA and that RNA could not have been the first molecule of life.

The cornerstone of the RNA world is the catalytic capability of a small ribozyme. Can a ribozyme molecule self-replicate naturally? Because ribozymes need to fold into specific complex structures to perform their functions, it is exceedingly unlikely that any self-replicating molecule can serve as its template. To serve as a template on which a new RNA molecule can be synthesized, a molecule must be unfolded and exposed to the monomers that will polymerize it. Moreover, unfolded molecules are not catalytic; catalysis is intimately related to the precise, three-dimensional configuration (Fig. 9.3c). To create synthetic ribozymes, scientists had to start from an already folded, catalytic RNA. They required 18–20 rounds of selection and optimization, which could not happen in the prebiotic world. Many challenges must be overcome to demonstrate that RNA could have supported an evolving genetic system by itself. Catalysis is a rare property of extremely long RNA sequences. It needs astronomical numbers (10^{14}–10^{16}) of randomized RNA molecules as a starting point for the isolation of the catalytic and binding functions of a synthetic ribozyme using in vitro selection. Compared to these numbers, our Milky Way contains only 10^8 stars, whereas the universe has 10^{21} stars. This is an awful lot of RNA molecules needed to isolate the designer ribozyme [34]. It is like choosing the right kind of star from a distant galaxy. Charles Carter of the University of North Carolina criticizes the above paper: 'The RNA world hypothesis is driven almost entirely by the flow of data from the extremely high technology of combinatorial libraries, whose relationship with the prebiotic world is anything but worthy of unanimous support' [40, 41].

The evolution of the RNA world would likely have reached a cul-de-sac because of the limited enzymatic functions of ribozymes. Critics point out that the catalytic repertoire of RNA is too limited. The preponderance of naturally occurring ribozymes catalyze phosphoryl transfer reactions; they make and break the RNA phosphodiester bond. Self-replicating RNAs may lack the fidelity needed to originate life. No one has achieved actual self-replication of RNA, which is the heart of the RNA world. Perhaps, RNA in isolation (including ribozymes) is simply not sufficient to catalyze its own replication, and substantial help from other molecules such as amino acids and peptides is essential. Amino acids and peptides are precursors to proteins. Evolution has resoundingly favored protein enzymes over ribozymes.

We propose that the self-replicator incorporated both peptides and RNAs. Such a peptide/RNA replicator is more feasible in the light of replication machinery currently found in cells than in RNAs alone. RNAs and peptides were critical molecules for the emergence of life-like systems in the prebiotic world. Recently, growing evidence has suggested that even at the early stages of the evolution of life, the functions of ribozymes would have been adequately assisted by abiotic peptides and other molecules present in their vent environment (see Fig. 8.2). Nearly all ribozymes found in living cells carry out their functions by association with proteins. These ribonucleoprotein complexes may represent remnants of ancient biology where peptides could have supported the folding and substrate binding of catalytic RNAs [35]. Before the advent of translation, ribozymes would acquire amino acids as 'cofactors' to increase their catalytic activity [36]. If RNA existed as a sole genetic biopolymer, it would have needed to replicate itself efficiently, completely, without peptides. It is a seductive picture—catalytic RNAs appeared by rare chance on early Earth as molecular replicators that gradually evolved into complex molecules capable of encoding proteins, metabolic systems, and, ultimately, DNA.

Studies of ribosomal evolution also challenge the RNA world hypothesis. A ribosome is a cellular complex of proteins and RNAs, which translates the genetic code. It is an ancient 'ribonucleoprotein machine' that formed as the result of a long and complex process involving the gradual

coevolution and accretion of RNAs and protein molecules before the origin of the first cells. The phylogeny of ribosomes, as discussed later, challenges the existence of the solitary RNA world. Ribosomal proteins must have coexisted with and contributed to ribosomes' formation, together with ribosomal RNA [24, 34]. A ribosome is fundamentally a peptidyl transferase ribozyme supported by proteins that help correct the RNA structure's folding and improve the efficiency and accuracy of the translation. Because many ribosomal proteins are evolutionarily conserved, it is challenging to imagine how ribosomes and translation could have evolved in a primitive RNA world without the assistance of proteins [33].

For all its elegance, the RNA world theory is premised on a significant assumption: that on primordial Earth, RNA could have copied itself. Although RNA can indeed drive some biological reactions, as far as we know, self-replication is not one of them. In recent years, an increasing number of researchers have become more supportive of an alternative view that now appears to be more realistic for explaining the origin of life—the peptide/RNA world.

10.5 Molecular Cooperation: An Alternative to the RNA World

The RNA world theory has been so alluring and expedient that many see no alternative. However, despite the conceptual sophistication of the RNA world, this hypothesis faces formidable difficulties in accounting for RNA synthesis and replication under plausible prebiotic conditions. In the prebiotic world, there were simple monomers like amino acids and nucleotides that eventually became polypeptides and RNAs. The RNA world might have appeared, but the theory could have overstated the exclusivity of RNA and neglected the role of peptides and lipid membranes. A primordial vent environment that could have supported RNA synthesis no doubt would have also spawned many other organic compounds—for example, peptide- and lipid-like molecules. There is no reason to believe that the vent environment exclusively produced a load of nucleotides or RNAs but discriminated against lipids and peptides for formation. Cosmic ingredients in the vent environment are more readily made amino acids and fatty acids, which can turn into peptides and lipids more easily than RNAs. The proto-metabolism arising in a Thioester world would give rise to amino acids and peptides [23]. The catalytic performance of these early simple peptides became critical in the various chemical reactions of RNAs.

Moreover, an RNA molecule is inherently fragile in the natural environment, degrading into smaller fragments through hydrolysis, thus preventing faithful replication. Perhaps peptides provided stability to RNA molecules. We see the compartmentalization of RNA and peptide molecules, leading to protocells as a crucial step toward systems that can undergo Darwinian evolution. As discussed earlier, abiotic vesicles may have also existed at the dawn of life. Carbonaceous meteorites have been found to contain amphiphilic molecules whose ability to bind to water and fatty compounds allow them to create a primitive membrane-like bilayer structure.

Additionally, regardless of whether the pioneer molecule was RNA or peptide, the 'single biopolymer paradox' must also be considered [38]. If RNA served for both genetic transmission and catalysis, these dual roles placed extremely different demands on it. Catalytic biopolymers must fold, whereas genetic biopolymers must not. Well-folded sequences are poor templates for copying, but poorly folded sequences are unlikely to be good ribozymes. Catalytic biopolymers appear to require many building blocks, whereas genetic biopolymers should have few. Catalytic biopolymers must catalyze reactions, whereas genetic biopolymers must not. In the RNA world scenario, catalytically active RNA enzymes would destroy RNA rather than create it. Parsimony would seem to favor a situation in which peptides support RNA functions.

A peptide/RNA world (frequently also called the RNA–peptide world) provides a division of labor that resolves the conflict between the templating activity required for RNA replication and the stable folding needed for catalysis by peptides. The RNA world violates the dual foundations of living systems—replication and metabolism [27, 37]. With no way to separate information storage from information processing, the RNA world violates the logical structure of life. In contrast, the coevolution of digital (RNA) and analog (peptide) systems in a peptide/protein world is more versatile and robust [38, 39]. Thus, the idea of an independent RNA world in the absence of peptides or amino acids as stabilizing structures that assist in catalysis is not a viable concept. Here, we propose instead a mechanism by which catalysis in proto-metabolic reaction networks could bring forth RNA as the dominant macromolecule for both catalysis and genetic information.

Charles Carter of the University of North Carolina developed the 'peptide/RNA' hypothesis, which suggests that life on Earth originated from an intimate partnership between RNAs and peptides and so contradicts the widely held RNA world hypothesis. He argues that a single polymer such as RNA could never perform all the necessary processes of living cells. Contrary to RNA's slow catalytic rate, Carter suggests that RNAs and peptides might have worked as complementary structures in the prebiotic environment to initiate translation and genetic coding [40, 41]. Carter and Wills have shown experimental results—a group of 20 ancient enzymes known as aminoacyl-transfer RNA (tRNA) synthetases (aaRSs)—that beautifully mesh with their peptide/

RNA theory [42]. Remnants of these ancient enzymes can still be found in all living cells and in some subcellular structures. These primordial enzyme superfamilies would have been formed along with the first genes and proteins. For Carter and Wills, RNA lacks the properties required for the spontaneous formation of life, mainly because it lacks a feature they call 'reflexivity.' According to this view, to eventually lead to life-forms, RNA needed peptides to form the necessary reflexive feedback loop, such that peptides and the nucleic acids encoding them would have mutually enabled each other's molecular organization.

The peptide/RNA world has received a boost from a recent experiment. Müller et al. [43] suggested that peptides and RNAs coevolved as a complex hybrid molecule, complementing each other's functions; they might have jointly served as a precursor to the modern ribosome. They found that the RNA molecules were helping the peptides grow, and the peptides were bringing stability to the RNA molecules. They provided a plausible pathway for how early noncoding RNA molecules (tRNAs and ribosomal RNAs) may have enabled peptides to grow directly on them, like mushrooms growing on a tree. These noncoding RNAs are considered living molecular fossils and were present in the last universal common ancestor (LUCA) of all organisms. This experiment suggests the early existence of a peptide/RNA world from which ribosomal peptide synthesis may have emerged.

10.6 Constructing a Peptide/RNA World

Whereas modern cells present no signs of a putative prebiotic RNA world, the possibility of coevolution between peptides and RNAs has received much attention in recent years. Inherent in these proposals is the idea that the functions of peptides and RNAs could have interacted in the prebiotic environment in ways subject to selection and chemical evolution. The emergence of the peptide/RNA world in the prebiotic environment is simple enough to have formed spontaneously yet complex enough to allow natural selection, leading to Darwinian evolution. Experiments reveal that supramolecular peptide structures bind and stabilize RNA [6]. Small peptides with essential amino acids can multiply the catalytic activities of ribozymes up to a hundred times. With distinct structures and functions, peptides and RNAs coevolving in the prebiotic world would have become codependent, working together to expand their informational, structural, and functional capabilities.

Molecular cooperation is a crucial factor in the origin of life. We endorse Carter's peptide/RNA world in elucidating the cooperative origins of the genetic code. We suggest that RNA first promoted reactions required for abiogenesis with the help of metals, amino acids, and peptides as cofactors. Then, as metabolism became more complex, RNA developed a translation system and created coded polypeptides. We have discussed the relevance of symbiosis for the origin of the genetic code: how a ribozyme would acquire an amino acid as a cofactor and a 'bridge peptide' would act as an activating enzyme. We argue for a mutualistic theory in which the different abilities of peptides and small RNAs reinforce each other. Still, abiogenesis quickly climbed the ladder of complexity [44].

In the peptide/RNA world, activated amino acids and peptides could have interacted with RNA molecules, possibly acting as cofactors for RNA replication and RNA-based catalytic reactions, thereby enhancing RNA stability, replication fitness, and ribozyme catalysis [36]. Moreover, interactions between amino acids or peptides and RNA genomes enhance their fidelity in copying genetic information, allowing the appearance of the first RNA genomes and the synthesis of numerous ribozymes with different enzymatic activities and leading to the emergence of new biosynthetic pathways for the biotic synthesis of peptides and polypeptides that predated the genetic code's appearance. Most likely, RNA replication was assisted by peptides and polypeptides before the emergence of biotic enzymes.

When assembled with peptides, for example, RNAs may have been better protected against hydrolysis. As more polymers joined the interactions, such chemical networks could have taken on increased functions. For example, short peptides with several acidic residues could have been important for protecting short RNA molecules against degradation by Mg^{+} ions [45]. Most likely, RNA and peptide molecules interacted very early on in the origin of the genetic code. Even short peptides had significant catalytic capabilities. A recent experiment has suggested that a ribozyme recruits an assortment of proteins to the RNA world as it evolves. Ribonuclease P (RNase P), one of the most ancient enzymes, is a ribozyme, unlike almost all others. This primordial ribozyme appears to co-opt proteins as it evolves. With an active site composed of RNA, it is present in living cells. It recognizes precursor transfer RNA (pre-tRNA), which processes mature tRNAs in collaboration with an assemblage of proteins, thus favoring the peptide–RNA world [46].

The three main species of RNA directly involved in protein synthesis are messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA). A ribosome is an RNA–protein complex performing protein synthesis, but the role of amino acids is overlooked in protein synthesis. The most crucial biochemical evidence for the antiquity of ribozymes comes from crystal structure determinations and other studies of the large subunit of the ribosome. The evidence confirms that the active site's peptidyl transferase is formed from domain V and 23S rRNA and does not involve proteins. However, this view is incomplete at best—the amino group of aminoacylated tRNA attacks peptidyl-tRNA, destroying the bond between the peptide and tRNA. Ribosomal proteins significantly contribute to

ribosomal function. The assembly of large and small subunits is dependent upon ribosomal proteins. Without ribosomal proteins, 23S rRNA is unable to serve as a peptidyl transferase. Several ribosomal proteins help in assembling the large subunit by providing unstructured, highly positively charged polypeptide sequences that bind RNA segments together and extend to the center of the subunit [46].

Perhaps there was a gradual addition of ribosomal proteins to ribosomal RNAs. It is possible that in the peptide–RNA world, ribosomal RNAs started to interact, first with small amino acids or peptides, enhancing the stability, efficiency, and specificity of the primitive ribosomes. At a later stage of evolution, genetically encoded proteins may have replaced these abiotic peptides. The putative ancient peptide segments from the ribosomal subunits' cores enhance RNA polymerase ribozyme function, as do derived peptides comprising lysine. An intricate relation between small RNAs and tiny peptides occurred in the emerging ribosomes, earliest synthetases, and translation systems. Prebiotic chemistry provides an increasingly clear indication that RNA replication arose not in isolation but in a chemically diverse environment, which included building blocks for simple peptides and lipids [47]. From unstructured peptide segments of several ribosomal proteins, short peptides can interact with and augment the function of RNA polymerase ribozyme and facilitate primordial RNA replication [49]. Overall, a ribosome, like RNase, is best described as a ribonucleopeptide catalyst that also requires Mg^2 for its function. The evolution of inherited genetic information coupled to encoded functional proteins is far more plausible than the solitary RNA world assisted by ribozymes sophisticated enough to give rise to a genetic code [24, 48].

However, there is still a broken thread between the nonenzymatic synthesis of RNA and the self-sustained replicases because of the lack of consideration of feedback processes. We hope to bridge the missing link by considering that RNAs and peptides coevolved in a symbiosis, complementing and reinforcing each polymer's chemical repertoire. RNA molecules would have played the role of translation, and catalytic peptides have been the prototypes of current polymerases. In summary, the peptide/RNA world appears to be the more parsimonious route to life than the lonesome RNA world [44, 50].

10.7 Conclusions

Among several competing hypotheses for the origin of life on early Earth, an RNA world, in which RNA assumed both informational and catalytic roles, has been widely accepted as the main paradigm. Compared with proteins and DNA, RNA is exceptionally versatile. Ribozymes serve as catalysts for chemical reactions between other RNA molecules. The discovery of catalytic ribozymes favored strong circumstantial evidence for the RNA world theory.

Despite its conceptual elegance, the RNA world hypothesis faces formidable difficulties, primarily the immense challenge of accounting for RNA synthesis under plausible prebiotic conditions. Although various building blocks of RNA molecules might have been delivered by carbonaceous chondrites, comets, and interplanetary dust particles such as sugar, phosphorous, and purine and pyrimidine nucleobases, a steady input of peptides would have been essential during the polymerization of activated nucleotides on the surface of clay substrates. Amino acids, contrariwise, could be easily polymerized on the mineral surface to form peptide molecules. An RNA molecule is inherently fragile in the natural environment and always degrades into smaller fragments through hydrolysis, preventing faithful reproduction. Peptides provided stability to RNA molecules. Additionally, the RNA world hypothesis has difficulty explaining the long RNA sequences needed for catalytic activity.

The RNA world is contested by the fact that a self-replicating molecule RNA polymerase ribozyme has yet to be discovered and that little evidence for its existence is seen in living world today. To isolate the catalytic and binding functions of a synthetic ribozyme, researchers using in vitro selection require astronomical numbers (10^{14}–10^{16}) of randomized RNA molecules.

The RNA world might have existed, but, in recent times, the RNA world paradigm has been shifting to a peptide/RNA world paradigm. A molecular replicator with two components—RNAs and peptides—overcomes these problems and may be a better fit. Most likely, RNA interacted with peptide molecules very early in the origin of the genetic code. RNAs and peptides coevolved as a complex hybrid molecule, and their intermingling might have sparked life. Even short peptides would have had significant catalytic capabilities. A ribozyme appears to recruit an assortment of proteins to the RNA world as it evolves [44]. We believe that in the prebiotic environment, RNAs and proteins were complementary molecules that formed a dynamic, mutually reinforcing system. Life depends on an assortment of interactive molecules; it is hard to imagine that it would have arisen from the dynamics of a single type of molecule. The differentiation between replication and metabolism intrinsic to living operations must have appeared before the origin of DNA. RNAs provide coded information to build proteins with the help of various enzymes. The establishment of the peptide–RNA mutualism was a crucial step in the origin of life. These two biopolymers with complementary functions became symbiotic, working in tandem to expand their informational, structural, and functional repertoires. In a peptide/RNA world, RNA could store genetic information, whereas peptide enzymes function as catalysts. The dual role of peptides

and RNAs in the prebiotic world is far more plausible than lonely RNA world scenario. The RNA world hypothesis cannot explain how the transition from RNA only to the protein/RNA world could have occurred.

References

1. Callahan MP, Smith KE, Cleaves SH, Ruzicka J, Stern JC, Glavid DP, Huse CH, Dworkin JP (2011) Carbonaceous meteorites contain a wide range of nucleobases. Proc Natl Acad Sci U S A 108:13995–13998
2. Furukawa Y, Chikaraishi Y, Ohkouchi N, Ogawa NO, Galvin DP, Dworkin JP, Abe C, Nakamura T (2019) Extraterrestrial ribose and other sugars in primitive meteorites. Proc Natl Acad Sci 116:24440–24446
3. Hazen RM (2006) Mineral surfaces and the prebiotic selection and organization of biomolecules. Am Mineral 91:1715–1729
4. Ferris JP, Ertem G (1992) Oligomerization of ribonucleotides on montmorillonite: reaction of the 5′-phosphorimidazolide of adenosine. Science 257:1387–1389
5. Rajamoni D, Vlassov A, Benner S, Coombs A, Olasagasti F, Deamer D (2008) Lipid-assisted synthesis of RNA-like polymers from mononucleotides. Orig Life Evol Biosph 38:57–74
6. Wieczorek R, Dorr M, Chotera A, Luisi PL, Monnard PA (2013) Formation of RNA phophodiester bond by histidine-containing dipeptides. ChemBioChem 14:217–223
7. Cech TR (2000) The ribosome is a ribozyme. Science 280:878–879
8. Robertson MP, Joyce GF (2012) The origins of the RNA world. Cold Spring Harb Perspect Biol 4:a003608
9. Woese CR (1967) The genetic code: the molecular basis of for genetic expression. Harper & Row, New York
10. Crick FH (1968) The origin of the genetic code. J Mol Biol 38:367–379
11. Orgel LE (1968) Evolution of genetic apparatus. J Mol Biol 38:381–393
12. Gilbert W (1986) The RNA world. Nature 319:616
13. Cech TR (1986) RNA as an enzyme. Sci Am 255(5):64–75
14. Altman S (2013) The RNA-protein world. RNA 19:589–590
15. Johnson WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. Science 292:1310–1325
16. Horning DP, Joyce GF (2016) Amplification of RNA by an RNA polymerase ribozyme. Proc Natl Acad Sci 113:9786–9791
17. Wochner A, Altwater J, Coulson A, Holliger (2011) Ribozyme-catalyzed transcription of an active ribozyme. Science 332:209–212
18. Zaher HS, Unrau PJ (2008) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. RNA 13:1017–1026
19. Wochner A, Atwater J, Coulson A, Holliger P (2011) Ribozyme-catalyzed transcription of an active ribozyme. Science 332:209–212
20. Tjhung KF, Shokhirev MN, Horning DP, Joyce GF (2020) An RNA polymerase ribozyme that synthesizes its own ancestor. Proc Natl Acad Sci U S A 117:2906–2913
21. Hazen RM (2005) Genesis: the scientific quest for life. Joseph Henry Press, Washington, DC
22. Wachterhäuser G (1993) The cradle of chemistry of life: on the origin of natural products in a pyrite-pulled chemoautotrophic origin of life. Pure Appl Chem 65:1343–1348
23. De Duve C (2005) Singularities: landmarks on the pathways of life. Cambridge University Press, New York
24. Harish A, Caetano-Anolles G (2001) Ribosomal history reveals origin of modern protein synthesis. PLoS One:e32776
25. Shapiro R (2007) A simpler origin of life. Sci Am 296:24–31
26. Trefil J, Morowitz HJ, Smith E (2009) The origin of life. Am Sci 97:206–213
27. Dyson F (2004) Origins of life. Cambridge University Press, Cambridge
28. Robertson DL, Joyce GF (1999) Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. Nature 344:467–468
29. Jaeger L, Wright MC, Joyce GF (1999) A complex ligase ribozyme evolved in vitro from a group of I ribozyme domain. Proc Natl Acad Sci 96:14712–14717
30. Orgel LE (2000) Self-organizing biochemical cycles. Proc Natl Acad Sci 97:12503–12507
31. Nielsen PE, Egholm M, Berg RH, Buchard O (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamids. Science 254:1497–1500
32. Cech TR (2009) Crawling out of the RNA world. Cell 136:599–602
33. Szostak JW, Bartel DP, Luigi P (2001) Synthesizing life. Nature 409:387–390
34. Bowman JC, Hud NV, Williams JD (2015) The ribosome challenge to the RNA world. J Mol Evol 80:143–161
35. Bernhardt HS (2012) The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). Biol Direct 7:23. https://doi.org/10.1186/1745-6150-7-23
36. Das C, Frankel AD (2003) Sequence and structure space of RNA-binding peptides. Biopolymers 70:80–85
37. Maynard Smith J, Szathmary E (1999) The origins of life. Oxford University Press, New York
38. Benner SA (2014) Paradoxes in the origin of life. Orig Life Evol Biosph 44:339–343
39. Walker SI, Davies PCW (2013) The algorithmic origins of life. J R Soc Interface 10:20120869. https://doi.org/10.1098/rsif.2012.0869
40. Carter CW Jr (2015) What RNA world? Why a peptide/RNA partnership merits renewed experimental attention. Life 5:294–320
41. Carter CW Jr (2016) An alternative to the RNA world. Nat Hist 125:28–33
42. Carter CW Jr, Wills PR (2017) Interdependence, reflexivity, impedance matching, and the evolution of genetic coding. Mol Evol. https://doi.org/10.1093/molbev/msx265
43. Müller F, Escobar L, Wu F, Wegrzyn E, Ninyte M, Amatov T, Chan CY, Pichler A, Carell T (2022) A prebiotically possible scenario of an RNA-peptide world. Nature 605:279–285
44. Chatterjee S, Yadav S (2019) The origin of prebiotic information system in the peptide/RNA world: a simulation model of the evolution of translation and the genetic code. Life 9:25. https://doi.org/10.3390/life9010025
45. Szostak J (2012) The eightfold path to non-enzymatic RNA replication. J Syst Chem 3. https://doi.org/10.1186/1759-2208-3-2
46. Lan P, Tan M, Zhang Y, Niu S et al (2018) Structural insight into precursor tRNA processing by yeast ribonuclease P. Science 362. https://doi.org/10.1126/science.aat6678
47. Carny O, Gazit E (2011) Creating prebiotic sanctuary: self-assembling supramolecular peptide structures bind and stabilize RNA. Orig Life Evol Biosph 41:121–132
48. Van der Gulik PTS, Speijer D (2015) How amino acids and peptides shaped the RNA world. Life 5:23–246
49. Tagami S, Attwater J, Holliger P (2017) Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function. Nat Chem 9:325–332
50. Kaddour H, Sahai N (2010) Synergism and mutualism in non-enzymatic RNA polymerization. Life 4:598–620

Noncoding RNAs: The Hybrid Information System

11

Life is more than just complex chemical reactions. The cell is also an information storing, processing, and replicating system. We need to explain the origin of this information and how the information processing machinery came to exist.

—Paul Davies, 2001

Although analog and digital information systems are well known in the literature, we identified a transitional form, the hybrid information system (HIS), on the basis of noncoding RNAs [1]. The close symbiosis between peptides and nucleic acids in modern cells indicates a functional coevolution between two polymers that led to the beginning of the peptide/RNA world. A molecular replicator with two components, namely, peptides and RNAs, played a critical role in building the hybrid information system, with a high degree of specified complexity. A hybrid information system makes use of information in both forms—the analog form and the digital form. It is analogous to a hybrid car powered by fuel in two forms—the gasoline form and the electricity form. The hybrid information system would enhance and perfect digital computing and give rise to the digital information system (DIS).

The age of hybrid information system arose as an emergent property of noncoding mRNA in the peptide/RNA world. It introduces molecular communication and complementarity—the symbiotic relationship—between RNAs and peptides. As discussed in the previous chapter, a molecular replicator with two components—peptides and RNAs—played a critical role in building the information system with a high degree of specified complexity.

RNA, as a single chain, is free to take any kind of shape. From the architecture of a single-stranded RNA molecule, different species of noncoding RNAs—such as ribozymes, transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs)—evolved inside protocells. Each species developed a supply of information, distinct in attribute and configuration, in response to the specific amino acids with which it interacted. The coordination of different species of RNA molecules signals the passage from the age of chemistry to the age of hybrid information. Noncoding RNAs would play critical roles in building the translation machine.

11.1 Noncoding RNAs

Coding RNAs generally refer to mRNAs that encode proteins—the latter acts as various components, including enzymes, cell structures, and mixed-signal transductors. mRNAs play key roles in the digital information system (DIS). Contrariwise, noncoding RNAs (ncRNAs) dominate the machineries for hybrid information system (HIS) . They can form abiotically in the vent environment by polymerization of nucleotides. ncRNAs belong to several groups and accomplish a variety of biological functions. The role of noncoding RNAs has gained increased attention in abiogenesis. Noncoding RNAs are not translated into proteins, but they are involved in making translation machines. Transfer RNAs (tRNAs) form an adaptor molecule between an mRNA and a protein. As discussed in Chap. 12, they would play critical roles in creating coding RNAs (mRNAs). A ribosome consists of more than 60% ribosomal RNA, which is made of three ncRNAs in bacteria.

The evolution of hybrid information systems during prebiotic synthesis must consist of reasonable elementary steps. Each step confers a distinct added advantage that leads to the efficient use of digital information. We highlight the salient features of the hybrid information systems during the major stages of abiogenesis. These stages are: (1) the emergence of noncoding RNA molecules such as ribozymes and tRNAs; (2) ribozyme–amino acid interaction and the origin of bridge peptides; and (3) the origin of ribosomes. Here, we outline the main features of these processes to highlight the evolution of hybrid information systems.

RNA is a versatile molecule with many variants and functions. It is free to take any kind of shape as a single chain. Because of its architectural flexibility, a single-stranded RNA molecule could give rise to different species of noncoding RNAs—such as transfer RNAs (tRNAs), ribozymes, and

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_11

ribosomal RNAs (rRNAs). Each species developed a unique configuration, attribute, and supply of information, in response to the specific amino acids with which it interacted. The coordination of different kinds of noncoding RNA molecules signals the passage from analog information to hybrid information. Here, we trace the origin and function of other species of noncoding RNAs in the peptide/RNA world. Many of these critical roles and functions of ncRNAs in the peptide/RNA world are believed to be molecular fossils, relics, or lost with the emergence of the first cells, and their current roles mainly remain in the regulation of information flow from DNA to proteins.

11.2 The Origin of Ribozymes

An RNA molecule has a secondary hairpin structure consisting of a double-stranded RNA stem by base pairing and a terminal loop (Fig. 11.1). The hairpin structure is the key to many secondary RNA structures, such as ribozymes and tRNAs. As an essential secondary structure of RNA, an RNA hairpin can direct RNA folding, determine interactions in a ribozyme, protect the structural stability of mRNA, provide recognition sites for RNA-binding proteins, and serve as a substrate for the enzymatic reaction [2]. Functional stabilization of ribozymes requires short peptide molecules, which were available in the peptide/RNA world. Structurally, RNA hairpins can occur in different positions within different types of RNAs; they differ in the length of the stem, the size of the loop, the number and size of the bulges, and the actual nucleotide sequence.

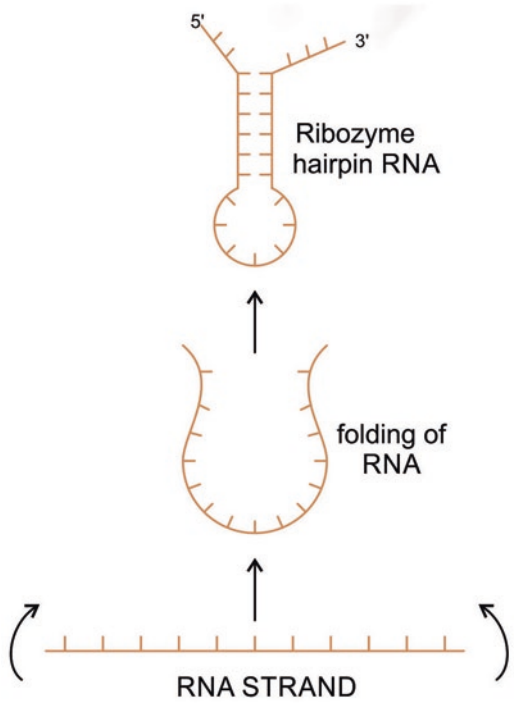


Fig. 11.1 The origin of the hairpin ribozyme and its chemical bonding with the appropriate amino acid and bridge peptide. A single-stranded RNA can develop a secondary structure by infolding with a double-stranded stem and a single-stranded loop, thus forming a hairpin ribozyme

Ribozymes are RNA molecules that can catalyze specific biochemical reactions in a similar way to protein enzymes. There are different classes of ribozymes, but they all appear to be associated with metal ions, such as potassium or magnesium, which are essential for the folding and catalysis of ribozymes. Different ribozymes catalyze different reactions, but almost all ribozymes catalyze the cleavage of RNA chains forming bonds between the RNA strands. However, the primary structure of RNA molecules is much more restricted than that of proteins by having only 4 bases versus the 20 types of amino acids at the base of proteins.

The chemical bonding of a particular amino acid with a small RNA hairpin structure led to the ribozyme's origin. We assume that different kinds of RNAs, enzymes nucleotides, oligonucleotides, and amino acids and peptides were available in the prebiotic soup. Owing to its single-stranded nature, an RNA molecule can be bent backward on itself in a hairpin loop, where the stems of the loop are maintained by base pairing to form a three-dimensional structure, just like a protein molecule, and act as an enzyme. In some stem–loop configurations, two ends of the stem might remain free, containing the 3′- and 5′-ends. The 3′-end might function as an acceptor stem to form a covalent bond with a specific amino acid. This small hairpin RNA molecule with specific terminal base sequences acquired the corresponding amino acid as a 'cofactor' to improve the catalytic range and efficiency to become initial ribozymes [3]. Many enzymes act with the help of one or more cofactors. The binding of amino acids to a ribozyme resulted in an enhancement of catalytic activity.

Ribozymes can also fold into various complex tertiary structures, i.e., three-dimensional structures that give ribozymes their catalytic ability (Fig. 11.2b). Many ribozymes configure either a hairpin- or hammerhead-shaped active center. All ribozymes catalyze the cleavage of RNA chains or the formation of bonds between RNA strands. The hairpin structure of a ribozyme is the key to many secondary RNA structures, such as tRNAs. An RNA template used complementarity in addition to protocellular information to enable base pairing and replication [2].

Ribozymes play a central role in harboring the HIS and creating translation machinery parts such as pre-tRNAs, bridge peptides, and ribosomes. In our previous work [4], we discussed the likely scenario for the origin of a pre-tRNA molecule from folded ribozymes with a stem-and-loop structure. The modern complex tRNA structure probably evolved from a simple precursor such as a pre-tRNA molecule (Fig. 11.3a–d). Two ribozyme molecules with hairpin

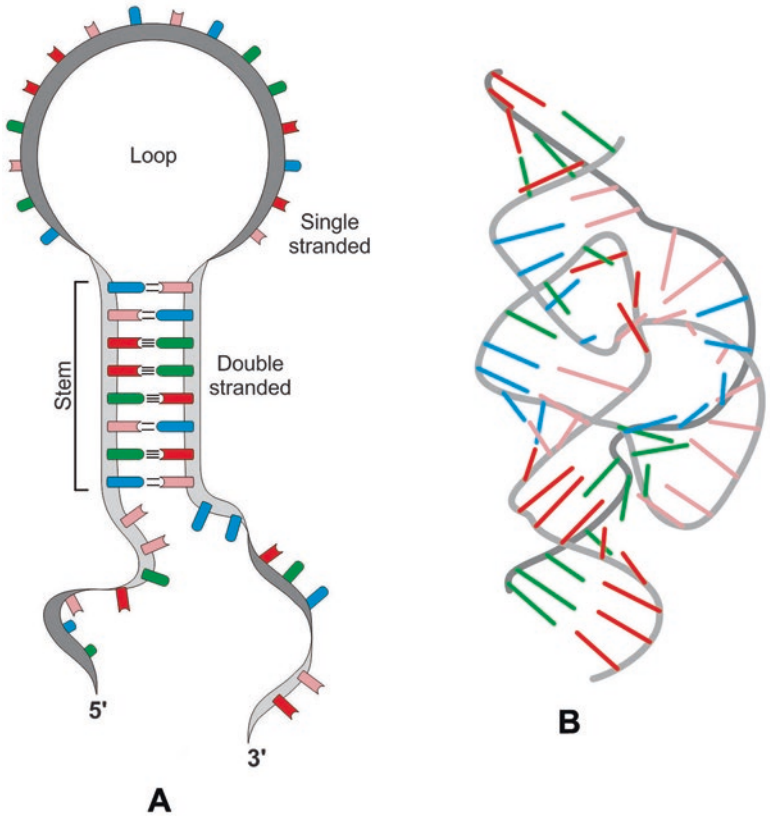


Fig. 11.2 (**a**) Although RNA is a single-stranded molecule, it can form a secondary hairpin structure of a ribozyme. (**b**) A hammerhead ribozyme can create a tertiary structure as in protein and catalyze reactions; the tertiary structure can have both Watson–Crick and noncanonical base pairs

structures with a stem and loop probably created a pre-tRNA molecule by fusion or ligation [5, 6].

Ribozymes also gave rise to a 'bridge peptide,' a precursor of pre-aaRS and aaRS (aminoacyl-tRNA synthetase) [4, 7]. Any specific binding between two molecules involves analog information functioning as if two molecules recognize each other. The ribozymes acquired an amino acid at its 3′-end as a cofactor; an amino acid was attached to a ribozyme and made it a more efficient catalyst [3]. Using cofactors, the range of catalytic activity can be increased (Fig. 11.3e).

Finally, ribozymes led to the origin of ribosomes (Fig. 11.3f). Within ribosomes, ribozymes function as part of the large subunit ribosomal RNA to link amino acids during protein synthesis. If ribozymes catalyzed certain reactions of amino acids and peptides, then selection of ribozymes would occur, which could bind specific amino acids and catalyze the synthesis of peptide bonds. Over time, those ribozyme activities would become the core catalytic function of ribosomes. A ribosome is fundamentally a peptidyl transferase ribozyme supported by ribosomal proteins (r-proteins) that contribute to the correct folding of the RNA structure and improve the efficiency and accuracy of translation [24]. These startup molecules of translation machinery derived from ribozymes would continue to evolve for efficiency and functionality, as discussed below.

11.3 The Origin of Transfer RNAs

Any model for the development of protein synthesis must necessarily start with direct interactions between RNAs and amino acids. Chemical considerations suggested that direct interactions between the amino acids and the codons in mRNAs were unlikely. The protein and mRNA languages seem to be unrelated. Amino acids do not read their codons. Some kind of adaptor molecule must mediate the specification of amino acids by codons in mRNAs during protein synthesis [9]. Other researchers soon identified the adaptor molecules as transfer RNAs (tRNAs), which serve as a reading device of mRNAs through base pairing. A tRNA molecule binds to amino acids, associates with mRNA molecules, and interacts with ribosomes to decipher and translate the mRNA code.

The first RNA gene, the *Ur-Gen*, was widely believed to be a precursor of modern tRNAs [10]. tRNA is an ancient molecule that has evolved very little over time. The phylogeny of ribosomes suggests that tRNAs are ancient components of ribosomes that arose in the early prebiotic world [11].

A tRNA molecule is short, typically 76–90 nucleotides in length. It serves as an adaptor molecule between mRNA and the amino acid sequences of proteins [12]. It conveys the information contained in the nucleotide sequences of mRNAs with the functional information contained in proteins.

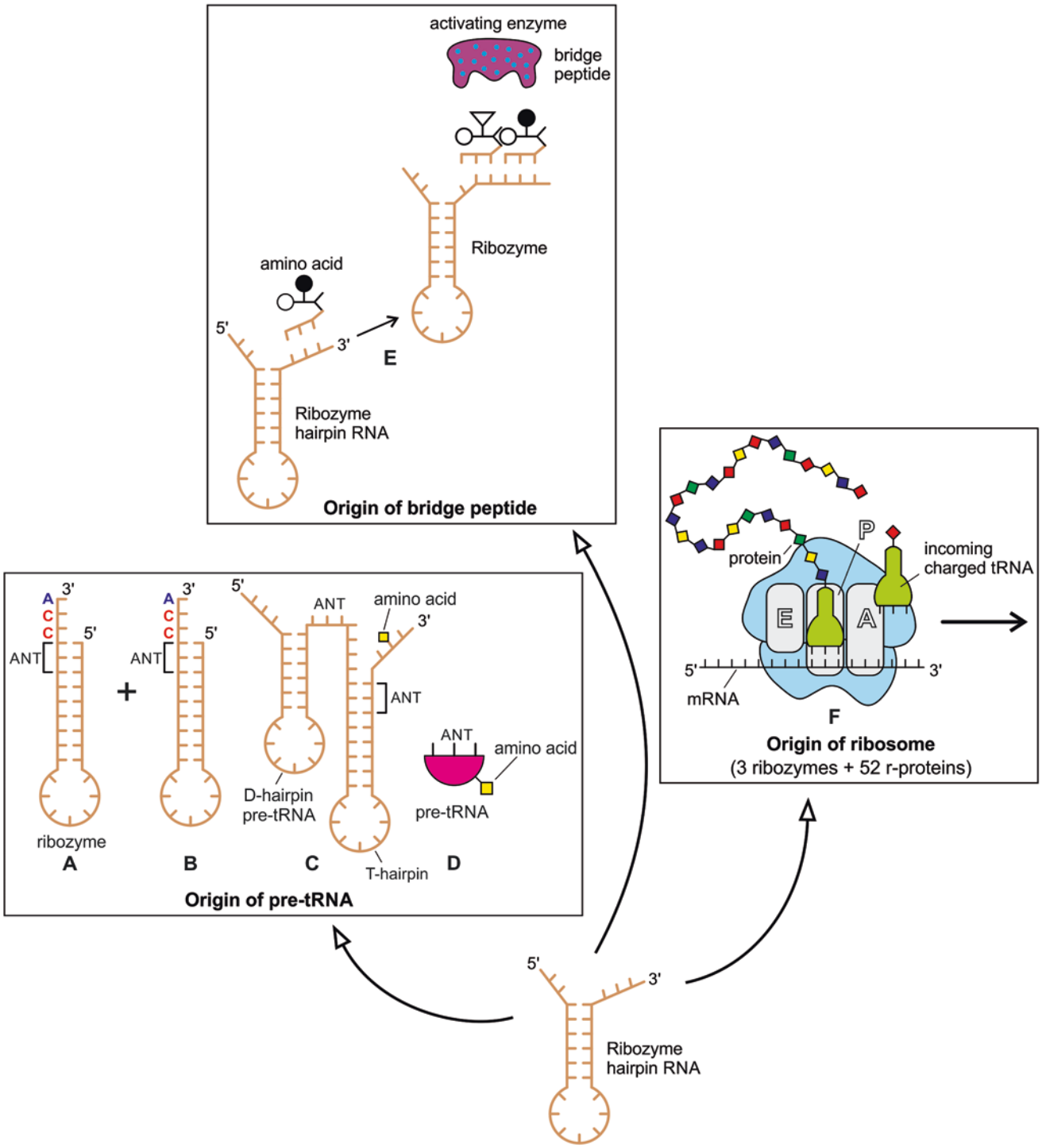


Fig. 11.3 The origin of three components of translation machinery with a stem and loop: pre-tRNA molecules (**a**–**d**), a bridge peptide (**e**), and a ribosome (**f**). (**a**–**b**) The hairpin structure of two ribozymes, each with a loop and a stem. (**c**) The ligation or duplication of the hairpin structures may give rise to a double-hairpin structure, forming a T-hairpin loop and a D-hairpin loop with an anticodon (ANT) site between the two stems. (**d**) A schematic and simplified diagram of the pre-tRNA molecule showing the anticodon site and amino acid attachment site. (**e**) The hairpin ribozyme structure with a stem and loop and its activating enzyme, the bridge peptide. The amino acid is attached to its free oligonucleotide end by the bridge peptide. (**f**) A ribosome, a hybrid ribonucleoprotein complex, decodes the message of mRNA to synthesize a small protein chain. It is a decoder of digital information to analog information

Although a tRNA molecule is short, both its primary structure and its overall geometry are undoubtedly more complex than those of any other RNA species [13]. The translation of a message carried in an mRNA into the amino acid language of proteins requires an interpreter. Converting the three-letter alphabet (codons) of nucleic acids to the one-letter alphabet of the amino acids of proteins, tRNA molecules serve as interpreters during translation. The amino acids themselves cannot recognize the codons in mRNA. The tRNA matches the appropriate amino acids to the appropriate codons. Each amino acid is joined to the correct tRNA by a special enzyme, namely, aminoacyl-tRNA synthetase (aaRS).

tRNAs can carry a specific amino acid and recognize the codon for that amino acid. It has a distinct amino acid attachment site on one end and an anticodon site at the opposite end. tRNA, therefore, functions as an adaptor molecule. tRNA participates in two distinct steps in the translation process. The first step comprises the reactions that lead to the charging of the tRNA molecule with an amino acid. The second step comprises the complex reactions in which the tRNA transfers its amino acids into a growing protein chain in response to a specific codon. The chemical reaction catalyzed by the tRNA is simple—the joining of amino acids through peptide linkages. It performs the remarkable task of choosing the appropriate amino acid to be added to the growing protein chain by reading successive mRNA codons. The actual step of translation from mRNA into protein language occurs when amino acids and tRNAs are matched and joined. The translators that do this job are the aminoacyl-tRNA synthetases (aaRSs). These enzymes are the only bilingual elements in a cell: they can recognize both the amino acid and the corresponding tRNAs. They are the critical elements of translation, being the links between the worlds of proteins and nucleic acids. The activation of tRNAs occurs when a synthetase uses energy from adenosine triphosphate (ATP) hydrolysis to attach an amino acid to a specific tRNA. There are 20 such synthetases, 1 for each amino acid. Together, they make up the complete dictionary for protein synthesis in a cryptic form that relies on tRNAs for decoding into the anticodon language. Each type of amino acid can be attached to only 1 type of tRNA, so each kind of organism has many types of tRNAs and more than 20 amino acids. There might have been a coevolutionary process in which the anticodons and the corresponding amino acids were progressively mediated by natural selection; as ribosomes appear, tRNAs transport amino acids to ribosomes, where they are assembled into proteins.

tRNAs would also create the first gene, as discussed in Chap. 12. Because of these critical roles, understanding the properties of tRNA molecules is critical in prebiotic information systems. Without tRNAs, genetics and coded protein synthesis are impossible. The primary structure and the overall geometry of tRNA molecules are undoubtedly more complex than those of any other RNA species.

The origin of tRNAs is contentious. Many studies have suggested that the modern cloverleaf structure of tRNAs may have arisen from a single ancestral gene by duplicating half-sized hairpin-like RNAs by passing through some intermediate structures such as pre-tRNA molecules (Fig.11.4a) [6, 13–16]. The linkage of a ribozyme to the amino acid at the terminal of a hairpin loop might be the starting point for the origin of tRNAs, a quarter the size of modern tRNA molecules. Pre-tRNA molecules, in turn, would give rise to tRNA molecules by structural rearrangement.

A tRNA molecule shows both secondary and tertiary structures (Fig. 11.4c, e). The secondary structure of a tRNA molecule in the solution with three hairpin loops resembles a cloverleaf from nature (Fig. 11.4c). One of these hairpin loops contains the anticodon, which forms base pairs with the codon of mRNA. The other two loops of the cloverleaf create a T-arm and a D-arm. The CCA sequence at the 3′-end of the acceptor stem forms a covalent bond with the amino acid that corresponds to the anticodon sequence. The CCA sequence of the acceptor stem offers a binding site for the amino acid. The 5′-terminal contains a phosphate group. The anticodon and the acceptor stem sequence correlate with amino acids' role in folded proteins [15, 16]. The secondary structure of a tRNA molecule may provide some clue as to its ancestral molecular configuration. The cloverleaf configuration of tRNA can be derived from a folded ribozyme with a single loop and an attachment site for the amino acid at the end of a stem.

The precise three-dimensional structure of tRNA molecules determines their function. The cloverleaf tRNA folds into an L-shaped tertiary structure, but each has a distinct anticodon and an attached amino acid (Fig. 11.4e). One arm of the L-shaped tRNA structure has a minihelix with a single-stranded CCA end that is used for connecting a single amino acid; the other arm forms an anticodon loop, with three unpaired bases that may bind to the complementary codon of mRNA. Each tRNA molecule can carry 1 of the 20 different amino acids at its CCA minihelix end. Each type of amino acid has its kind of tRNA, which binds it and brings it to the growing end of a protein chain during the decoding of mRNA. The CCA end of the minihelix interacts with the large ribosomal subunit to form a peptide bond, and the loop end communicates with the small ribosomal subunit for decoding mRNA triplets through codon–anticodon interactions [14].

Aminoacylation of tRNA is an essential event in the translation system. In the modern system, protein enzymes play the sole role in tRNA aminoacylation; in the primitive translation system, ribozymes could have catalyzed aminoacylation to tRNAs or ancestral tRNA-like molecules. What

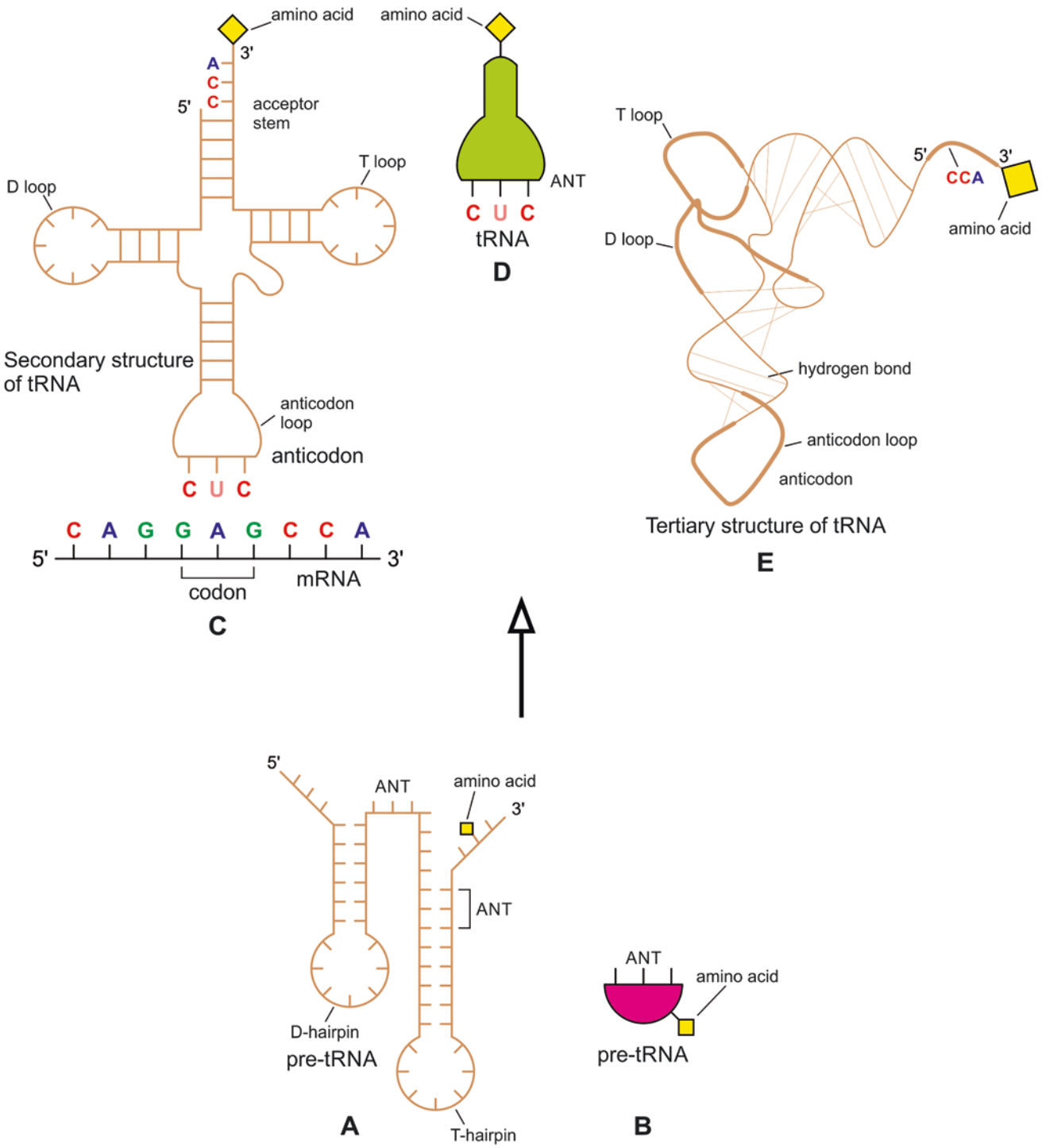


Fig. 11.4 (**a**) The evolution of a tRNA molecule from a precursor pre-tRNA molecule (**a**, **b**) by gene duplication. (**c**) The secondary structure of a tRNA molecule could have been created by ligation of two half-sized pre-tRNA structures. Now, a full-length tRNA structure looks like a cloverleaf; its anticodon end forms a complementary base pair with the codon of mRNA. (**d**) A simplified and schematic diagram of the tRNA molecule showing the site of the anticodon. (**e**) The cloverleaf secondary structure of tRNA could be folded into an L-shaped tertiary structure; it shows the aminoacylation site at the CCA end. The minihelix region (half domain of the tRNA with the amino acid attachment site) interacts with the conserved domain of aaRS for amino acid activation. The other half of tRNA interacts with the non-conserved domain of aaRS for specific recognition of an anticodon. (Modified from Chatterjee and Yadav [1, 4])

was the catalytic function of a ribozyme? If it attaches an amino acid to its end, then it would not be logical that the substrate amino acid is the cofactor simultaneously. This attachment first occurred to make cofactors, and ribozymes carried it. The RNA world hypothesis implies that the ribozyme functioned as an assignment enzyme to attach a particular amino acid to an ancestral tRNA for aminoacylation before the emergence of aaRS [17]. In the peptide/RNA world, we suggest that the ribozyme was not an aminoacylation catalyst; another molecule, such as a bridge peptide, performed this function for the ligation of the amino acid with ancestral tRNA [7]. In the early stage of aminoacylation, pre-aaRS, originally a protein enzyme, emerged as an assignment enzyme for charging ancestral tRNA [8, 18, 19]. In that case, the ribozyme should perform another function that is so advantageous as to help the molecule survive. In our view, the ribozyme's cofactor function was utilized to form peptide bonds between adjacent amino acids before the ribosome's emergence. This enzymatic activity may be a precursor to that of the peptidyl transferase center (PTC) of the ribosome responsible for the peptide bond formation. Another phenomenon in which a ribozyme's intervention could have been of critical importance is RNA replication [12].

The most plausible scenario for the origin of a tRNA molecule is based on ribozymes. The chemical bonding of specific amino acids with small RNA molecules with specific base sequences was the crucial step. Perhaps the precursor of tRNA started as a simple ribozyme with a hairpin structure (Fig. 11.4). This ribozyme acquired amino acids at its 3′-end as a cofactor: that is, an amino acid was attached to a ribozyme and made it a more efficient catalyst [3]. By using cofactors, the range of specificity of catalytic activity could be increased. One way of attaching an amino acid to a particular point on the ribozyme's surface is to attach it to the end of a single-stranded unlooped stem of the hairpin, which is charged and begins to bind the amino acid, thus enhancing the catalytic function of the ribozyme. With the stabilization of catalytic reactions, these ribozymes started to participate in the first catalytic cycles. This configuration of a ribozyme linking an amino acid to the end may be the starting point for the origin of tRNAs, where the unlooped stems contain the free 3′- and 5′-ends of the chain. This amino acid attachment to ribozymes by a specific assignment enzyme first occurred to make cofactors more efficient catalysts [3].

The relevance of ribozymes in the origin of tRNAs is enormous. The equivalent effect of gene duplication might be accomplished by a simple ligation of two identical hairpins of folded ribozymes to create double hairpins, a D-hairpin and a T-hairpin, with an anticodon at the stem bases [15]. RNA ligation is a powerful driving force behind the emergence of tRNA, joining two hairpin loops of a ribozyme (Fig. 11.4c). During the evolutionary transitions of the pre-tRNA molecule, the double-hairpin structure with the D-hairpin and the T-hairpin formed in the ancient prebiotic world anticodon in the terminal CCA sequence adjacent to the D-hairpin (Fig. 11.4d) [20].

We suggest that this half-sized hairpin structure of the pre-tRNA molecule acquired some functional capacity for translation before the emergence of tRNAs (Fig. 11.4a, b). The pre-tRNA molecule is the evolutionary precursor of the tRNA molecule. Direct duplication or ligation of half-sized, hairpin-like structures—the pre-tRNA molecule— could have formed the contemporary full-length tRNA molecules (Fig. 11.4c, d). The acceptor stem bases and the anticodon stem/loop bases in tRNA in tRNA 5′-half and 3′-half fit together with the double-hairpin folding; this suggests that the primordial double-hairpin RNA molecules could have evolved to the structure of modern tRNAs by gene duplication, with subsequent mutations to form the familiar overleaf structure [4, 20]. In other words, two pre-tRNA molecules somehow fused to form a tRNA molecule.

The half-sized pre-tRNA molecule with two loops (D-hairpin and T-hairpin) on one side, and the anticodon and acceptor stem region of the CCA end on the other side, is structurally and functionally independent and is more ancient than the other half of the tRNA molecule [14]. This short, self-structured strand of the pre-tRNA molecule possesses a template domain, which is chargeable through interaction with specific amino acids and is probably the predecessor of tRNA (Fig. 11.4a, b). This pre-tRNA molecule binds, with high specificity, to the amino acid corresponding to its anticodon; this reaction is catalyzed by a specific pre-aminoacyl-tRNA synthetase (pre-aaRS). The tRNA evolution is closely linked to aminoacylation. There is a separate tRNA for each amino acid that carries a triplet sequence of nucleotides for anticodon. Later, the anticodon of pre-tRNA will guide the codon formation of the pre-mRNA. The pre-tRNA and tRNA molecules became bilingual that recognized both the 4-letter alphabet of nucleic acids and the 20-letter alphabet of amino acids. A ribozyme had developed HIS in the form of a molecular distributed hybrid information system to use mechanisms such as RNA splicing, RNA cleavage, and peptide synthesis in various reactions.

11.4 The Origin of Bridge Peptide, Pre-aaRS, and aaRS Enzymes

Aminoacyl-tRNA synthetases (aaRSs) are a superfamily of enzymes that are responsible for creating the pool of correctly charged aminoacyl-tRNAs necessary for the translation

of the genetic information (mRNA) through ribosomes. aaRSs are extremely ancient enzymes that are present in all organisms and are one of the pioneer molecules that are formed by the polymerization of amino acids in incremental steps. Each enzyme catalyzes the activation of specific amino acids and recognizes a specific tRNA for binding. These enzymes thus implement the genetic code because they act as molecular dictionaries that can read both the 3-letter code of mRNAs and the 20-letter code of amino acids.

The aminoacylated ribozyme was the pioneer molecule to use an amino acid as a cofactor and employed a bridge peptide for activation. Later, with the development of tRNAs, the activation reaction is catalyzed by specific aaRS, a derived product of a bridge peptide. The first step is the formation of an aminoacyl adenylate with amino acid and ATP. The next step is to transfer the aminoacyl group to a particular tRNA molecule to form an aminoacyl-tRNA or a charged tRNA. The mechanism of aaRS formation is well-known [5]. It reveals how and why the tRNA molecule creates the bilingual enzyme aaRS that connects it with the appropriate amino acid. It enhances the selection and sorting of the proper amino acids from the prebiotic soup for protein synthesis. Each aaRS is highly specific for a given amino acid. It has a highly discriminating amino acid activation site. Both amino acids and ATP were available in hydrothermal vents, facilitating a reaction with tRNA to form aminoacyl-tRNA synthetase. Moreover, the proofreading ability by aaRS increases the fidelity of protein synthesis.

How do aaRSs choose their tRNA partners? The aaRSs recognize, on the one hand, individual amino acids, which they activate via conjunction with adenosine triphosphate (ATP), or the aaRSs activate amino acids to generate their conjugate with adenosine monophosphate (AMP) [17]. The synthetase first binds ATP to the corresponding amino acid to form an aminoacyl-adenylate, releasing inorganic pyrophosphate (PP_1). The next step is to transfer the aminoacyl group of aminoacyl-AMP to a particular tRNA molecule to form aminoacyl-tRNA. The mechanism can be summarized in the following reaction series:

1. Amino acid + ATP → Aminoacyl-AMP + PP_1
2. Aminoacyl-AMP + tRNA → Aminoacyl-tRNA + AMP

Thus, the equivalent of two molecules of ATP is consumed in the synthesis of each

aminoacyl-tRNA. One is consumed in the formation of the ester linkage of aminoacyl-tRNA, whereas the other is consumed in driving the reaction forward. The same aminoacyl-tRNA synthetase catalyzes the activation and transfer steps for a particular amino acid. Indeed, the aminoacyl-AMP intermediate does not dissociate from the synthetase. Aminoacyl-AMP is normally a transient intermediate in the synthesis of aminoacyl-tRNA. Synthetases can recognize the anticodon loops and acceptor stems of tRNA molecules. Their precise recognition of tRNAs is as vital for high-fidelity protein synthesis as is the accurate selection of amino acids. Besides, the synthetases perform a 'proofreading' or 'editing' activity that can remove mischarged amino acids from the enzyme or tRNA molecule.

aaRSs come in 20 variants, with each one being specific to an amino acid and a tRNA. These 20 enzymes are widely different, with each being optimized to function with its particular amino acid and the set tRNA molecules appropriate to that amino acid. They can be divided into two classes, termed 'class I' and 'class II'—the two aaRS superfamilies evenly split translation into 10 amino acids each. The initial activating enzyme was a bridge peptide that facilitated the aminoacylation of the ribozyme (Fig. 11.5). From a bridge peptide, protozymes, then urzymes, and, finally, pre-aaRSs and aaRSs probably evolved [7, 18, 19, 21]. We speculate that the precursor of aaRS was pre-aaRS, a hypothetical primordial ancestor that gave rise to two classes of aaRSs, which are both multidomain proteins. Each aaRS uses different mechanisms of aminoacylation. In our model, the original aminoacylation enzymes were pre-aaRS, a simpler version of aaRS, which must have featured a strong linkage to a pre-tRNA molecule anticodon. This linkage must have featured a codon-like, trinucleotide-binding site for the adaptor's anticodon on the pre-aaRS.

We propose that pre-aaRS is an enzyme, including an anticodon, plus a domain capable of binding and activating an amino acid and transferring it to the pre-tRNA. Pre-aaRS is analogous to 'protozymes' and 'urzymes' [18, 19] but is somewhat more advanced because it would allow for tRNA/anticodon recognition. Protozymes retain about 40% of the activity of the full-length of aaRS, even though they contain only about 10% as many amino acids. Next came 'urzymes,' which retain about 60% of the activity and have the same functional repertoire as the full-length enzymes. We speculate that pre-aaRS would be as long as the urzyme but that it acquired an additional anticodon-binding function. The proposed evolutionary path from bridge peptide to protozyme to urzyme to pre-aaRS to aaRS documents increases the complexity would satisfy the rule of continuity [7].

11.5 The Origin of Ribosomes

Translation needs one more component of the molecular machine to organize proteins in a continuous assembly line—the ribosomes. Ribosomes are ancient molecular machines that consist of ribosomal proteins and ribosomal RNAs. They read the nucleotide sequence of an mRNA into a protein sequence using the genetic code. They use tRNAs to mediate the translation process from the nucleotide language of mRNAs into the amino acid language of proteins.

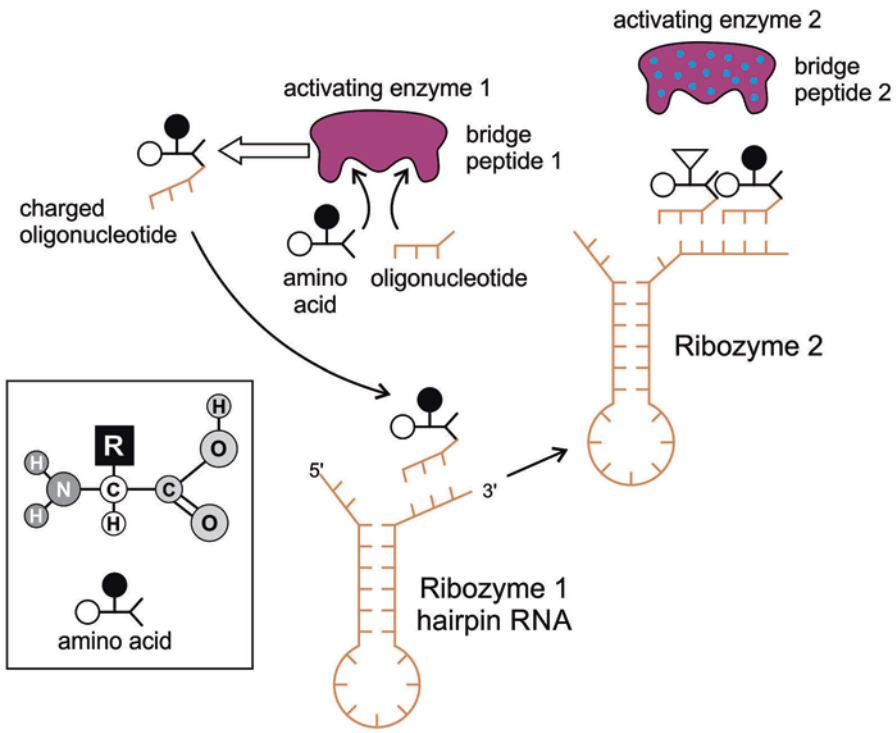


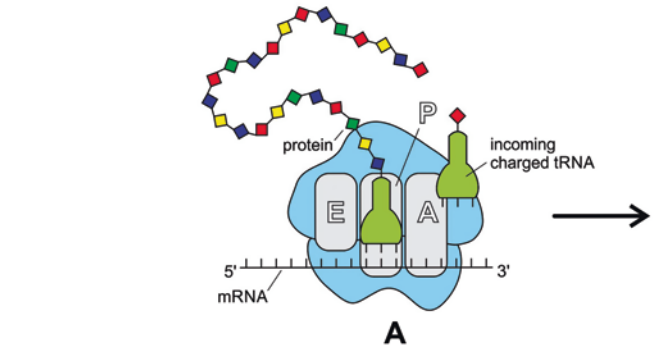
Fig. 11.5 The hairpin ribozyme structure with a stem and loop and its activating enzyme, the bridge peptide. The amino acid is attached to its free oligonucleotide end by the activating enzyme 1, such as a bridge peptide (ribozyme 1). The activating enzyme 2 joins the next batch of amino acids, and the oligonucleotide is attached to ribozyme 2, forming the peptide bond. Inset, an amino acid has the same general structure: a central carbon bonded with an amino group, a unique side chain or R-group, and a carboxyl group. (Modified from Chatterjee and Yadav [4])

Cells typically contain many thousands of ribosomes in the cytoplasm region. Each ribosome can bind one mRNA and up to three tRNAs. A ribosome develops three binding sites for tRNA molecules—the A, P, and E sites (Fig. 11.6a).

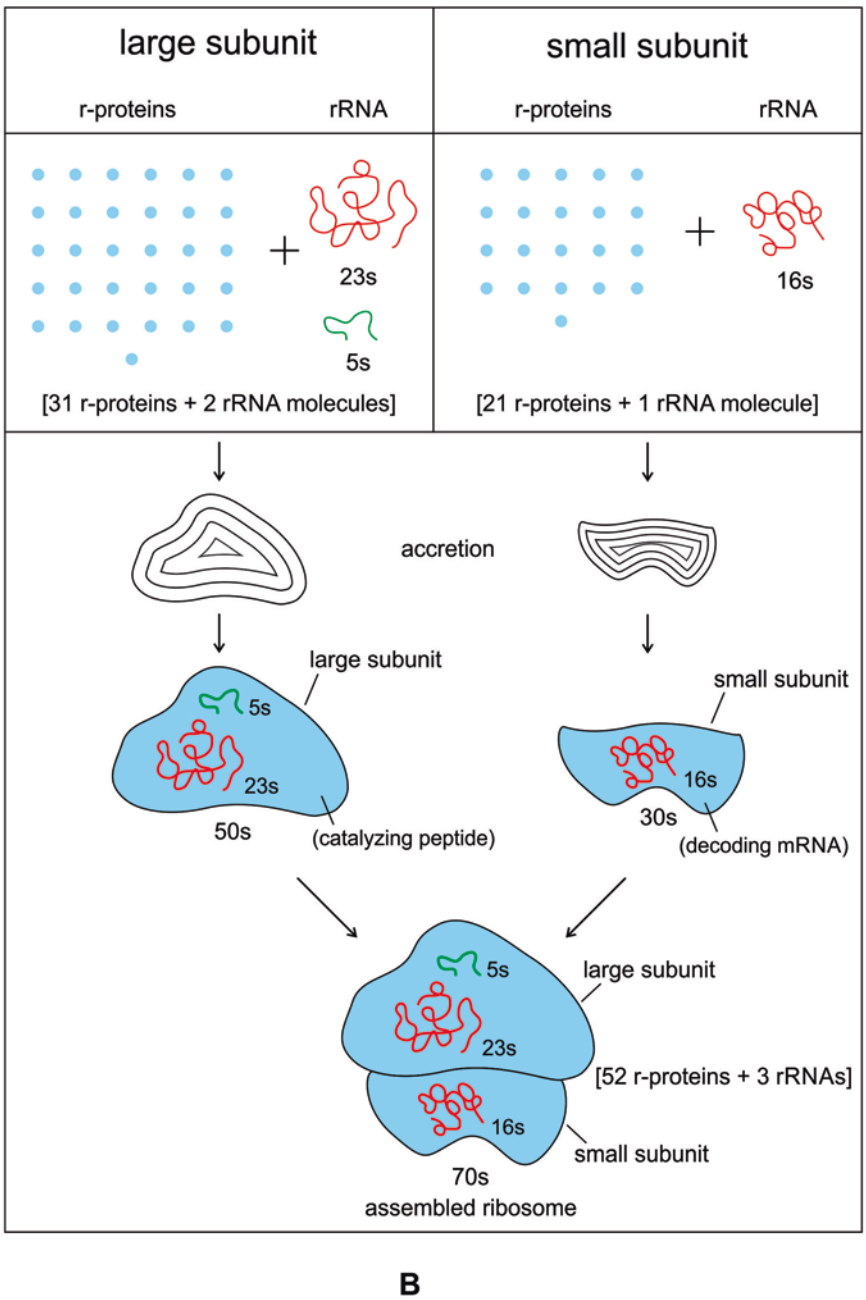
Ribosomes link amino acids together in the order that is specified by mRNA molecules. They provide the environment for controlling the interaction between the codons of mRNA and anticodons of aminoacyl-tRNA in the creation of proteins. The translation of encoded information of mRNA and the linking of amino acids that were selected by tRNAs are at the heart of the protein production process. Ribosomes can link amino acids together at a rate of 200/min. Therefore, small proteins can be made quickly. Once a new protein chain is manufactured, the ribosome is released from protein synthesis to enter a pool of free ribosomes in equilibrium with separate small and large subunits [22].

A ribosome is composed of two-thirds of RNA and one-third protein. It is made of about 50 ribosomal proteins (r-proteins) that are wrapped up with 4 ribosomal RNAs (rRNAs), and, it is, therefore, a ribonucleoprotein (RNP) (Fig. 11.6). Although ribosomal proteins greatly outnumber ribosomal RNAs, the rRNAs account for more than half of the ribosome's mass. A bacterial cell may contain as many as 20,000 ribosomal complexes, which enable the continuous production of several thousand different proteins, both to replace degraded proteins and to make new ones for daughter cells during cell division. A ribosome physically moves along an mRNA strand, reads the codon sequences of the mRNA, and catalyzes the assembly of amino acids into protein chains using the genetic code. It uses tRNAs to mediate the translation process from the nucleotide language of mRNAs into the amino acid language of proteins with the help of various accessory molecules. Each ribosome can bind one mRNA and up to three tRNAs. Central to the development of ribosomes are RNAs that spawn the tRNAs, and a symmetrical region deep within the large ribosomal RNA, where the peptidyl transferase reaction occurs [5, 23, 24].

Recent bacterial ribosomes shed light on the origin, evolution, morphology, and composition of primitive ribosomes that emerged in the peptide/RNA world. The bacteria have smaller ribosomes, termed '70S ribosomes,' which are composed of two major subunits of unequal size, which are called the large (50S) and the small (30S) subunits; each consists of one or two RNA chains and scores of proteins (Fig. 11.6). The small subunit (SSU) is where mRNA and tRNA molecules interact with reading the genetic code, and the large subunit (LSU) is where the growing protein chain is synthesized from the amino acids that are attached to tRNAs. Thus, the small subunit mainly decodes mRNA, but the large subunit has mostly a catalytic function. In the large subunit, rRNA performs an enzyme's function, and it is termed a 'ribozyme.' In prokaryotic ribosomes, the small subunit, the 30S, is made of 1 ribosomal RNA and 21 ribosomal proteins, whereas the large subunit, 50S, is made of 2 ribosomal RNAs and 31 ribosomal proteins. The two subunits fit snugly into a slot, with a strand of the mRNA molecule running between



A



B

Fig. 11.6 A ribosome. (**a**) The translation machinery of a ribosome decodes the message of mRNA to synthesize a small protein chain. It is a decoder of digital information to analog information. During translation, the ribosome moves along the mRNA chain from left to right in the 5′- to 3′-direction. (**b**) The ribosome is a molecular hybrid of two polymers, a ribonucleoprotein complex of 3 rRNA and 52 r-protein molecules. The ribosome consists of two subunits, large (the 50S) and small (30S); the small subunit holds mRNA during gliding and decodes the message of amino acids, and the large subunit links amino acids by catalyzing peptide bonds to create a protein chain. Most likely, the symbiotic interactions between ribosomal RNAs and ribosomal proteins gave rise to ribosomes, which grew by accretion. However, there is some controversy about whether the small or large subunits appeared first. In our view, both units coevolved by the accretion of ribosomal RNAs and ribosomal proteins. (Modified from Chatterjee and Yadav [1, 4])

them, similar to a tape inside a cassette player. The ribosome glides through the mRNA tape to read the encoded message in codons, which then carries out its instructions bit by bit, linking the amino acids together in a specified sequence until a complete protein has been synthesized. Ribosomal RNAs are programmed to recognize the codon as it appears on mRNA. When specific protein production is finished, the two subunits of the ribosome drift apart [23, 24]. Ribosomes only have a temporary existence. The large and small subunits of a ribosome undergo a cycle of association and dissociation during each round of translation. Similarly, once the protein is made, mRNA is broken down, and the nucleotides are recycled.

Ribosomes evolved prior to the emergence of DNA and cellular life in the peptide/RNA world. Ribosomal evolution is intricately linked to the prior evolution of mRNAs, tRNAs, and the primitive form of the genetic code and translation. The origins and evolution of ribosomes remain printed in the extant life's biochemistry and the ribosomal structure. Most theories propose that the ribosome was a functional takeover of a primitive RNA-based translation system in a coordinated series of chemical reactions. RNAs are believed to be responsible for the bulk of the ribosome's work. Recent structures of ribosomes have unambiguously shown that the essential functions of a ribosome, such as decoding, peptidyl transfer, and translocation, appear to be mediated by RNAs [25, 26]. The crystal structure of a ribosome suggests that it is fundamentally a peptidyl transferase ribozyme, supported by r-proteins that contribute to the correct folding of the RNA structure and improve the efficiency and accuracy of translation. The phylogeny of ribosomes suggests that the origin of rRNA is linked to accretionary tRNA building blocks that gave rise to functional rRNA [11]. The decoding center where mRNA is in the small subunit is primarily formed from 16S rRNA. rRNAs are folded into highly compact and precise three-dimensional structures that form the core of ribosomes. rRNAs give ribosomes their overall shape. Thus, the widely popular concept of 'the ribosome is a ribozyme' was born; the ribozymes must have preceded coded protein synthesis [27].

In recent times, the role of proteins in the origin of ribosomes has been gaining currency, implying that ribosomes may have first originated in a peptide/RNA world, where both amino acids and a variety of peptides were available [10, 12, 16, 17, 28, 29]. Ribosomal proteins are not passive contributors to ribosomal function. They are generally located on the surface, where they fill the gaps and crevices of the folded rRNAs. The main role of ribosomal proteins is to fold and stabilize the rRNA core while permitting the changes in rRNA conformation necessary for this RNA to catalyze efficient protein synthesis. Ribosomal proteins provide the structural framework for the 23S rRNA, which carries out the peptidyl transferase reaction. In the absence of ribosomal proteins, 23S rRNA is unable to exhibit peptidyl transferase activity: the assembly of large and small subunits that are dependent upon ribosomal proteins [30, 31]. Several ribosomal proteins assist in assembling the large subunit by providing unstructured, highly positively charged protein sequences that bind amino RNA segments together and extend to the center of the subunit [30]. These extensions cooperatively fold with ribosomal proteins to produce the small subunit.

Why would an RNA structure evolve to make proteins if the protein did not already exist that would confer a selective advantage on the ribosomes capable of synthesizing them? The availability of even simple proteins could have significantly enlarged the otherwise limited catalytic function of RNAs. Many prebiotic protein enzymes carried out several key functions in the primitive translation system. Moreover, the production of simple proteins had already commenced through the interactions between tRNAs and aaRSs before the origin of ribosomes (Fig. 11.1). Perhaps ribosomal proteins were synthesized in the primitive translation system, which was then recruited to build the ribosome step by step. RNAs and proteins developed a symbiotic relationship to create ribosomes in the peptide/RNA world [11–13]. These r-proteins took an active part in stabilizing the evolving ribosomes and in interacting with many rRNA sequences. Because the number of proteins greatly exceeded the number of RNA domains, it can hardly be a surprise that every rRNA domain interacts with multiple proteins in ribosomes [42]. Ribosomes are not entirely ribozymes but are more accurately ribonucleoprotein (RNP), a complex that can have as many as 62 r-proteins, with only 3 rRNA molecules (Fig. 11.6). Primordial ribosomes could have benefited from even rather short basic peptides produced by random condensation of primitive amino acids such as glycine, alanine, aspartic acids, and valine, available in the hydrothermal crater vent environment.

Virtually, all r-proteins are in contact with rRNAs. Accordingly, it makes sense that this assemblage is a result of a long and complicated process of gradual coevolution of rRNAs and r-proteins. Both the assembly and synthesis of the ribosomal components must occur in a highly coordinated manner [14]. Their phylogenetic analysis reveals that the ribosomal protein/rRNA coevolution manifested throughout the prebiotic synthesis process, but the oldest proteins (S12, S17, S9, L3) appeared together with the oldest rRNA substructures that were responsible for both the decoding and ribosomal dynamics (3.3–3.4 Ga). Although different RNA molecules mostly carry out protein synthesis within the ribosome, such as mRNA, tRNA, rRNA, and peptidyl transferase, aminoacyl synthetase (aaRS) played a crucial role in a protein enzyme that attached the appropriate amino acid to its tRNA during protein synthesis. In terms of importance, the synthetases are equal to the tRNAs in the decoding

process because it is their combined action that allows each codon in the mRNA molecule to associate with its proper amino acid.

Similarly, both rRNA and the 50S subunit proteins are necessary for the peptidyl transferase activity during peptide bond formation. Still, the actual act of catalysis is a property of the ribosomal RNA of the larger subunit (Fig. 11.6). With the emergence of the primitive translation process, r-proteins began to emerge for building ribosomes. Ribosomes, in turn, would make the translation machine sophisticated and efficient.

The accretion model describes the origin and evolution of ribosomes [14]. Given that ribosomes are quite ancient, rRNAs and r-proteins likely coevolved to build this complex nanomachine. Like the tree rings, the ribosome contains a record of its history, spanning 4 billions of life on earth. They accreted to grow bigger and bigger over time. However, after they accreted, the older parts froze, like the rings of a tree (Fig. 11.6). Recent phylogenetic works on ribosomal history have suggested that both RNAs and proteins contributed to forming the ribosomal core through accretion, recursively adding expanding segments [14, 15]. Ribosomes contain life's most ancient and abundant polymers, the oldest fragments of RNAs and protein molecules. It is most likely a molecular relic of the peptide/RNA world [11].

Both ribosomal subunits have separate functions. Peptide bond formation occurs at the peptidyl transferase center (PTC) of the large subunit, whereas the mRNA sequences are decoded on the small subunit. mRNA decoding contributes to the specificity of protein synthesis on the ribosome. In isolation, both of the subunits can perform their respective functions (Fig. 11.6). By itself, the large subunit will catalyze the formation of peptide bonds between aminoacyl-tRNA-like substrates. By itself, the small subunit binds mRNA, and, when mRNA is bound, it binds tRNAs in a codon-specific manner. In an RNA world scenario, ribosomes originated at the peptidyl transferase center of the large ribosomal subunit [24, 25]. There are no r-proteins that are close to the reaction site for protein synthesis. This suggests that the ribosomes' protein components do not directly participate in the peptide bond formation catalysis, but rather the proteins act as a scaffold that may enhance the ability of rRNAs to synthesize proteins. Ribosomes themselves, although being fundamentally ribozymes in nature, still require r-proteins to fold their rRNAs into biologically active conformations and to optimize the speed and accuracy of their functions. The ribosomal surface is an integrated patchwork of rRNAs and r-proteins. Ribosomes created the symbiosis between nucleic acids and proteins.

Currently, there is a debate regarding the origin of ribosomal subunits: which unit came first, the small or the large subunit? The PTC of large ribosomal subunits likely evolved from pre-tRNA molecules by duplicating the minihelix [20]. In this view, the simple function of peptide bond formation at the PTC site came first, and the specifications based on the codon sequence came later. In other words, the large subunit of the ribosome came first, followed by the addition of the small unit. However, these proposals do not link protein synthesis to RNA recognition and do not use a comparative phylogenetic framework to study ribosomal evolution.

Another author, who, from the ribosome's phylogeny, deduced that the small unit of the ribosome came first, offers a contrasting view of the origin of ribosomal subunits [14]. This study suggests that the small ribosomal subunit components evolved earlier than the catalytic peptidyl transferase center of the large ribosomal subunit. In this view, ribosomal RNAs and proteins coevolved tightly, starting with the oldest proteins (S12 and S17) and the oldest rRNA helix in the small subunit (the ribosomal ratchet responsible for ribosomal dynamics), ending with the modern multi-subunit ribosomes. A significant transition in the evolution of ribosomes at around 4 Ga brought independently evolving subunits together by infolding the inter-subunit contacts and interaction with full cloverleaf tRNA structures.

In our view, both the small and large subunits of the ribosome simultaneously appeared and worked in tandem [4]. Both decoding of mRNA and peptide bond formation were essential components during protein synthesis. These two subunits might have coevolved to join during translation and separate after protein synthesis. rRNAs are folded into highly compact, precise three-dimensional structures to form the core of a ribosome, whereas the r-proteins are generally located on the surface, where they fill the gaps and crevices of the folded RNA and act to fold and stabilize the core [32]. As these two subunits expanded through accretion, eventually arriving at the size of the bacterial ribosome, the accretion stopped; they then bound together during protein synthesis and finally split apart when the ribosome finished reading its mRNA molecule (Fig. 11.6).

If the fundamental functions of a ribosome are based on rRNAs, then why are there so many ribosomal proteins, some of which are highly conserved? One explanation is the rRNAs do not fold into their functional state in the absence of r-proteins. Another reason for the presence of proteins in ribosomes is that they improve the efficiency and accuracy of the translation [33]. Both rRNAs and r-proteins combine to perform the multitask procedure of protein synthesis in ribosomes. Harish and Caetano-Anolles suggested that a ribosome's functionally important and conserved regions were recruited and could be relics of an ancient peptide/RNA world [11]. Most likely, the biosynthetic mechanism responsible for primordial peptides and ancient r-proteins must have existed, which was superseded by ribosomes over time.

According to this accretionary model, very early in ribosomal evolution, rRNA helices interacted with r-proteins to

progressively form a core that mediated nucleotide interactions, which later served as the center for the coordinated and balanced RNP (ribonucleoprotein) accretion that evolved into our modern ribosomal function [14]. The early existence of smaller functional units of the ribosome, which can carry out different translational steps, such as peptidyl transferase, decoding, and aminoacylation, along with the development of A, P, and E sites for the positioning of tRNA molecules, can be inferred from the phylogeny. These small functional RNA/protein units were incrementally accreted and then refined by incorporating additional rRNA and r-protein molecules. Similarly, the first atomic resolution of the larger of the two subunits of the ribosome suggests that the large subunit's RNA components accomplish the key peptidyl transferase reaction [32]. Thus, rRNA does not exist as a framework to organize catalytic proteins. Instead, proteins are structural units, and they help hold the critical ribozyme. A 'pure' RNA world is incompatible with the existence of the coevolutionary pattern that is proposed for ribosomal molecules.

Perhaps rRNAs, such as noncoding ribozymes, acquired amino acids as cofactors, making them more efficient catalysts. By using cofactors, the range and specificity of catalytic activity can be increased. Ribozymes would have been in greater need of cofactors than protein enzymes because, without them, the range of reactions that they can catalyze is much smaller [6].

In our endosymbiotic model, rRNAs and r-proteins were brought into proximity within the phospholipid membrane to form the building blocks of the primordial ribosome. The origin of the ribosome precursor, through fusion and accretion of these ribosomal RNA and protein molecules' key components, is the likely scenario. rRNA and r-protein molecules began to fuse because of a chiral preference and then formed rudimentary ribosomes. Once the ribosome's core formed, the mRNA and tRNA molecules were recruited to help in translation through a trial-and-error method. Once the right mRNA and the small core subunit of ribosomes were in place, the ribosome would become increasingly complex by adding early conserved rRNA and r-proteins. Ribosomal proteins played an important role in supporting the ribosomal structure and in promoting translation. With the onset of operational coding, tRNAs began to assemble amino acids into long chains of proteins. Here, we suggest that a ribosome-like entity was one of the critical intermediates between prebiotic and cellular evolution, which is formed by endosymbiosis and the fusion of rRNA and r-protein molecules. Once ribosomes were installed inside the protocell membranes, the translation system was much improved.

In vitro constructions of ribosomes can shed new light on the mechanism of protein synthesis and provide deeper insights into the way that nature has assembled this complex machine. Working with *Escherichia coli* cells, natural ribosomal proteins were combined with synthetically made rRNAs, which self-assembled in vitro to create semisynthetic, functional ribosomes [34, 35]. Comprising 57 parts—3 strands of rRNAs and 54 proteins—an artificial ribosome (termed 'Ribo-T'), in which 2 subunits are tethered together by a short length of RNA, can carry out standard translation and pump out custom-made proteins. The ability to make ribosomes in vitro is a process that mimics nature and opens new avenues for the study of ribosomal synthesis, suggesting the coevolution of ribosomal RNAs and proteins.

Most likely, two separate functions of ribosomes evolved simultaneously by accretion growth, the decoding functions in the small subunit and the peptidyl transferase center in the large subunit. The availability of simple proteins could have significantly enlarged the otherwise limited catalytic functions of the ribozyme. The ribosome might have first originated as a ribozyme that only later evolved for structural complexity when ribosomal proteins began to appear in primitive translation. These r-proteins stabilized the structure and complexity of evolving ribosomes and interacted with many rRNA sequences. Both the assembly and synthesis of ribosomal components must occur in a highly coordinated symbiotic system [4].

A ribosome manifests the beautiful cooperation between two polymers—RNAs and proteins. A ribosome is a hybrid machine composed of one-third protein and two-thirds RNA. In this elegant machine, about 50 ribosomal proteins (r-proteins) are wrapped up with 4 ribosomal RNAs (rRNAs). A ribosome is, therefore, a ribonucleoprotein (Fig. 11.6b). The rRNAs contribute to more than half of the ribosome's mass.

Ribosomal RNAs are programmed to recognize the pairing between mRNA codons and complementary tRNA anticodons to a growing polypeptide chain (Fig. 11.6a). When specific protein production is complete, the two subunits of the ribosome become separated and are recycled after each round of translation. Similarly, once the protein is made, mRNA is broken down, and the nucleotides are recycled. A ribosome is a complex hybrid nanobot that integrates analog and hybrid information systems and communicates between them.

11.6 Components of the Translation Machine

Although a translation machine would develop step by step in response to translating the encoded mRNA strand for protein synthesis, its major components were already developed at this hybrid information stage. These machinery parts include the precursor of tRNA, such as pre-tRNA, ancestral

aminoacyl-tRNA synthetase, such as pre-aaRS, and, finally, the ribosomes. In the digital information system (DIS), these components of the translation machine would coevolve with three stages in mRNA in a molecular orchestra to translate the digital information content of mRNA into the analog format of the protein.

11.7 Conclusions

Noncoding RNAs such as ribozymes acquired a new kind of information system called hybrid information system (HIS). Hybrid information systems started to take advantage of digital computing with qualities such as accuracy, memory, accurate replication, and faithful transmission by combining it with analog computing. Every biological system is made up of two types of interconnected parts—an analog part that processes analog quantities and interacts with the world and a digital part that processes digital quantities and helps preserve and reproduce itself. Ribosomal RNAs are mainly hybrid components, but they function as catalysts like proteins. Ribozymes are critical molecules encoding HIS that gave rise to pre-tRNAs, pre-aaRSs, and, finally, ribosomes. Ribosomes perform two critical functions in prebiotic synthesis: (1) translate encoded information in mRNA to amino acids (2) and link together amino acids into a polypeptide chain. The output of a ribosome is the synthesis of the protein chain, which is a three-dimensional analog information system. Thus, a ribosome can be considered an efficient digital-to-analog converter.

We have discussed three critical molecular components of the translation machine: pre-tRNAs, pre-aaRSs, and, finally, ribosomes, which would work in a complex repertoire to decode the digital information embedded in the nucleotide sequences of mRNA for the synthesis of proteins. All the components of translation machines are bilingual and orchestrated nicely like an elaborate factory production line during the manufacturing of proteins. Once the translation machinery was in place, digital information could enter the system.

References

1. Chatterjee S, Yadav S (2022) The coevolution of biomolecules and prebiotic information systems in the origin of life: a visualization model for assembling the first gene. Life 12:834. https://doi.org/10.3390/life12060834
2. Chen IA (2015) Prebiotic chemistry: Replicating towards complexity. Complexity:101–102
3. Maynard Smith J, Szathmary E (1999) The origins of life. Oxford University Press, New York
4. Chatterjee S, Yadav S (2019) The origin of prebiotic information system in the peptide/RNA world: a simulation model of the evolution of translation and the genetic code. Life 9:25. https://doi.org/10.3390/life9010025
5. Freeman S (2005) Biological science, 2nd edn. Pearson Prentice Hall, Upper Saddle River
6. Di Giulio M (1992) On the origin of transfer RNA molecule. J Theor Biol 159:199–209
7. Kunnev D, Gospodinov A (2018) Possible emergence of sequence specific RNA aminoacylation via peptide intermediary to initiate Darwinian evolution and code through origin of life. Life 8. https://doi.org/10.3390/life804044
8. Noller HF (1993) On the origin of ribosome: coevolution of subdomains of tRNA and rRNA. In: The RNA World. Cold Spring Harbor Laboratory Press, Plainview, pp 137–156
9. Crick FHC (1970) Central Dogma in molecular biology. Nature 227:561–563
10. Eigen M, Winkler-Oswatitsch R (1981) Transfer-RNA, an early gene. Naturwissen 68:282–292
11. Harish A, Caetano-Anolles G (2001) Ribosomal history reveals origin of modern protein synthesis. PLoS One. https://doi.org/10.1371/journal.pone.0032776
12. De Duve C (2005) Singularities: landmarks on the pathways of life. Cambridge University Press, New York
13. Woese CR (1965) On the evolution of the genetic code. Proc Natl Acad Sci 54:1546–1552
14. Di Giulio M (2004) The origin of tRNA molecule: implications for the origin of protein synthesis. J Theor Biol 226:89–93
15. Di Giulio M (1995) Was it an ancient gene codifying for a hairpin RNA that, by means of direct duplication, gave rise to the primitive tRNA molecule? J Theor Biol 177:95–101
16. Tamura K (2015) Origins and early evolution of the tRNA molecule. Life 5:1687–1699
17. Van der Gulik PTS, Speijer D (2015) How amino acids and peptides shaped the RNA world. Life 5:23–246
18. Carter CW Jr (2016) An alternative to the RNA world. Nat Hist 125:28–33
19. Carter CW Jr, Wills PR (2017) Interdependence, reflexivity, impedance matching, and the evolution of genetic coding. Mol Biol Evol. https://doi.org/10.1093/molbev/msx265
20. Tanaka T, Kikuchi Y (2001) Origin of cloverleaf shape of transfer RNA-the double hairpin model: implication for the role of tRNA intro and the long extra loop. Viva Org 29:134–142
21. Carter CW Jr (2015) What RNA world? Why a peptide/RNA partnership merits renewed experimental attention. Life 5:294–320. https://doi.org/10.3390/life5010294
22. Wong JTF (1975) A coevolution theory of the genetic code. Proc Natl Acad Sci U S A 72:1909–1912
23. Ikehara K (2002) Origins of gene, genetic code, protein and life: comprehensive view of life systems from a GNC-SNS primitive code hypothesis. J Biosci 27:165–186
24. Noller HF (2004) The driving force for molecular evolution of translation. RNA 10:1833–1837
25. Noller HF (2012) Evolution of protein synthesis from an RNA world. Cold Spring Harb Spec Biol 4:a003681
26. Ramakrishnan V (2010) The ribosome: some hard facts about its structure and hot air about its evolution. Cold Spring Harb Symp Quant Biol 74:155–179
27. Cech TR (2000) The ribosome is a ribozyme. Science 280:878–879
28. Damer B, Deamer DW (2015) Coupled phases and combinatorial selection in fluctuating hydrothermal pools: a scenario to guide experimental approaches to the origin of cellular life. Life 5:872–887
29. Martin WF, Sousa FL, Lane N (2014) Energy at life's origin. Science 344:1092–1093
30. Altstein AD (2015) The progene hypothesis: the nucleoprotein world and how life began. Biol Direct 10:67. https://doi.org/10.1186/s13062-015-0096-z
31. Osawa S (1995) Evolution of the genetic code. Oxford University Press, Oxford

32. Schimmel P, Alexander RW (1998) All you need is RNA. Science 28:658–659
33. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. Science 289:920–929
34. Orelle C, Carlson ED, Szal T, Florin T, Jewett MC, Mankin AS (2015) Protein synthesis by ribosomes with tethered subunits. Nature 524:119–124
35. Savir Y, Tlusty T (2013) The ribosome as an optimal decoder: a lesson in molecular recognition. Cell 153:471–479

The First Gene Before DNA: The Digital Revolution

12

Genes are pure information—information that can be encoded, recoded and decoded, without any degradation or change of meaning. Pure information can be copied and since it is a digital information, the fidelity of copying can be immense.

—Richard Dawkins, 1995

12.1 Aminoacyl-tRNA

Today, digital information is stored in DNA and mRNA. Messenger RNA (or mRNA) is created from DNA transcription to direct protein synthesis and to perform other essential functions of life that DNA cannot do. Here, we explore how mRNA was designed step by step by a tRNA molecule before the rise of DNA during prebiotic synthesis. The aminoacyl-tRNA synthetases (aaRSs) are ancient and a universally distributed family of enzymes—the lynchpin between the worlds of peptide/RNA and proteins. Each aaRS is the matchmaker between a tRNA and an amino acid by selecting the correct amino acid from the prebiotic soup and pairing it with its cognate tRNA, corresponding to its anticodon (see Sect. 11.5 on aaRS). aaRS needs to discriminate the tRNA molecule as well, to ensure correct coupling of amino acid and tRNA. This happens based on the anticodon and the acceptor stem of the tRNA, being recognized by the aaRS anticodon-binding domain and the catalytic domain. For each of the canonical amino acids, there is a different aaRS and one or more tRNAs. However, aaRS never imparted the knowledge of the amino acid–anticodon relationship to tRNA. A tRNA always depends on its aaRS partner for selecting and catalyzing appropriate amino acids. aaRS becomes a personal catalyst for each tRNA in prebiotic synthesis. A tRNA never learned how to pick up and attach an appropriate amino acid at its 3′-terminal. An uncharged tRNA had limited function at this stage of abiogenesis.

An aminoacyl-tRNA (aa-tRNA or a charged tRNA) has acquired two different kinds of information systems at its two ends: first, it created an empty codon strand or template by the older anticodon base pairing system and, second, it transferred or encoded the transient amino acid information to the codon. The goal of aa-tRNA was to build a separate device where it could transfer these two sets of information permanently without any degradation. This is a logical step in abiogenesis. aa-tRNAs became the molecular architects for assembling mRNA codons to store their amino acid information. Each encoded codon is a separate linear nucleotide sequence in which the digital information of amino acids is encoded (Fig. 12.1a). Initially, four amino acids were available in the prebiotic soup for making four kinds of charged tRNA molecules. These charged molecules created innumerable encoded codons by combinations. These codons were joined to polycodons in digital sequences. As new amino acids became available, new kinds of charged tRNA molecules would appear to create new kinds of codons. Eventually, polycodons are linked to longer codon chains to form pre-mRNA and mRNA molecules.

Walker defined molecular memory as a specific mapping mechanism or as any correspondence between the input and the resulting output of an event [1]. A charged tRNA molecule not only created a codon by base pairing its anticodon but also permanently transferred the secret of appropriate amino acid information to the codon. We call this novel mechanism the ‘memory transfer model’ (MTM) from aminoacyl-tRNA to a new emerging gene. aa-tRNA transfers not only amino acids to ribosomes during protein synthesis but also its cognate amino acid information during the creation of mRNA. Thus, the use of ‘transfer’ in naming this aa-tRNA is doubly meaningful. The input from aa-tRNA determines the coding assignment for mRNA as an output.

Our memory transfer model may solve one of the long-standing riddles concerning the coding assignments in nucleic acids: which codons encode for which amino acids. How did these designations come about? As nucleic acid bases and amino acids do not recognize each other directly but can only do so via aa-tRNAs, there is no apparent reason why particular codons should encode specific amino acids. Once mRNA is assembled, aa-tRNA could read the mRNA

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_12

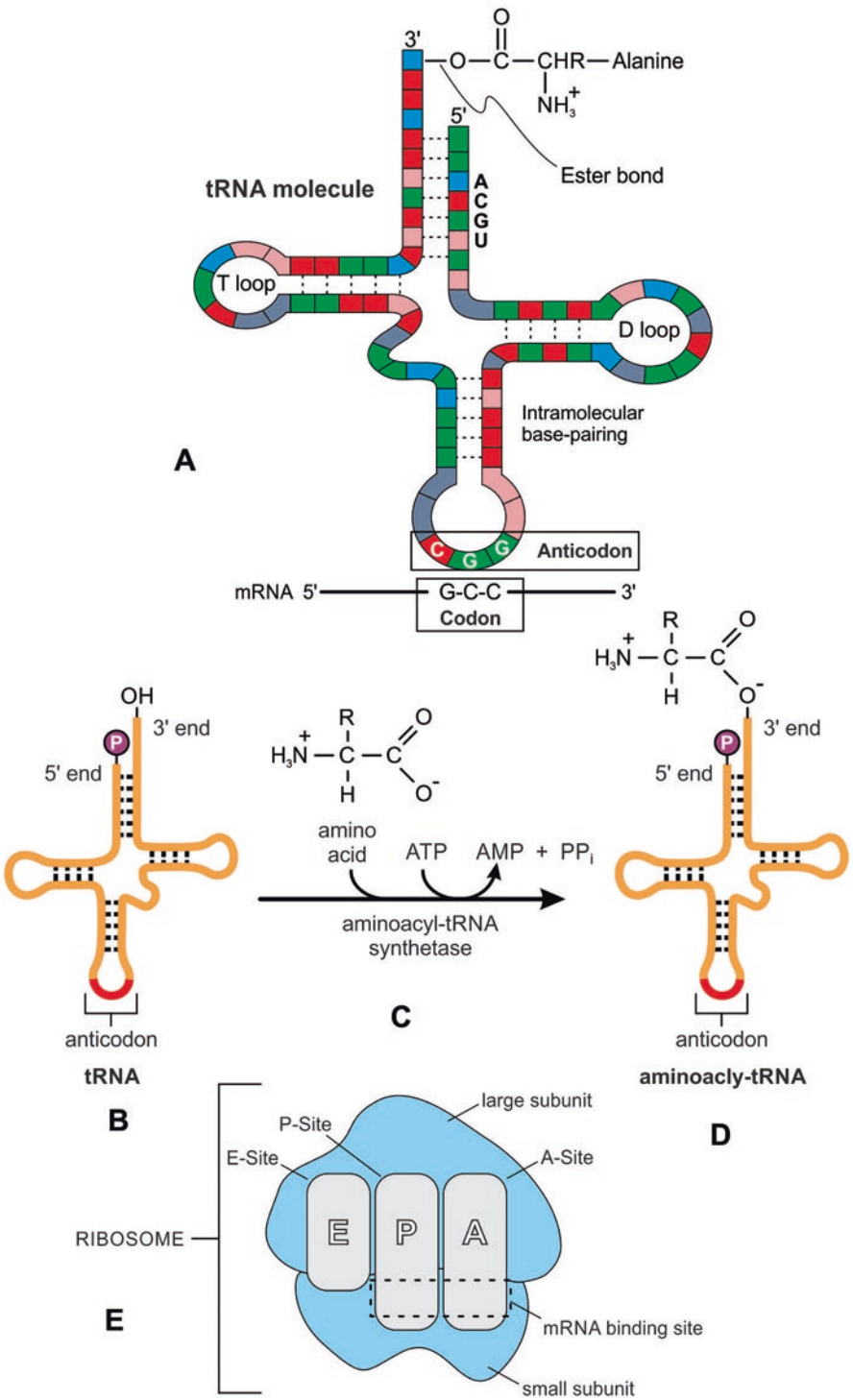


Fig. 12.1 Critical components of the translation machine. (**a**) Cloverleaf secondary structure of an aminoacyl-tRNA molecule shows two information centers at two opposing ends that would be responsible for creating mRNA by transfer of information. At one end, its cognate amino acid is attached; the opposite end shows antiparallel binding of the anticodon to the mRNA codon for alanine (GCC). (**b**) A tRNA shows a 5′-terminal for the phosphate group and a 3′-terminal for amino acid attachment. (**c**) A two-step aminoacylation reaction by aaRS (aminoacyl-tRNA synthetase); in the first step, aaRS binds ATP and its corresponding amino acid to form aminoacyl-adenylate (AMP), releasing inorganic pyrophosphate (PP_i); the second step involves the transfer of AMP to tRNA. (**d**) Aminoacyl-tRNA with its cognate amino acid. (**e**) A ribosome showing two subunits, i.e., a large ribosomal subunit and a small ribosomal subunit; each ribosome has a binding site for mRNA and three binding sites for tRNA. *E* **e**xit of deacylated tRNA, *P* most recent amino acid, **p**eptidyl tRNA, *A* incoming **a**minoacyl tRNA

codon for protein synthesis. The interaction between aaRS and its cognate tRNA bridged the transition from the hybrid information system (HIS) to the digital information system (DIS). The age of DIS began with the emergence of mRNA. Storage and directed transfer of mRNA information was the key requirement for the origin of life. aa-tRNA molecules were the unsung hero behind the creation of the first gene before the origin of DNA.

12.2 The Late Digital Revolution

In the peptide/RNA world, an mRNA molecule encodes the information of amino acids in linear nucleotide sequences for making a specific protein. This was the beginning of the digital revolution. The function of mRNA as a generator of proteins is extremely old, perhaps prior to DNA itself. Like DNA, mRNA can store and replicate genetic information.

Without mRNA, no life supported by genetic coding could evolve. RNA's versatility is one reason that scientists believe that this polymer came first, with DNA evolving later to store genetic information permanently. DNA has greater stability and durability than does mRNA for storing digital information. However, how did mRNA or the first gene emerge before DNA? The origin of mRNA is poorly understood. mRNA is not a product of random assemblage from the prebiotic molecular milieu but is created precisely with encoding information in its codon sequences. Here, we offer a plausible biochemical pathway for the evolution of the first gene step by step by tRNA [2].

With the sequential evolution of a variety of noncoding RNA molecules in the hydrothermal crater vent environment, such as ribozymes, pre-tRNAs and tRNAs, pre-aaRSs and aaRSs, and, finally, ribosomes—creating the different parts of a translation machine—the next logical biochemical step would be the emergence of an mRNA-like molecule to coordinate these noncoding RNAs in a network needed for translation of the digital encoding information (Fig. 12.1). Complexity of contingency and directionality in prebiotic synthesis led to the emergence of mRNA.

The passage from the hybrid information system (HIS) to the digital information system (DIS) was a smooth transition in the peptide/RNA world. The HIS facilitated the emergence of hierarchical translation machines, which helped coevolve the genome content of mRNA molecules in a feedback loop. Barbieri [3] suggested that life is artifact-making and that digital information plays a significant role in manufacturing artifacts. In the peptide/RNA world, the DIS in mRNA helped manufacture proteins by translation machines.

Digital information in life has the ability to store and process the information necessary for its communication and reproduction. It is characterized and sustained by several information-rich biological processes that govern cellular functions and significantly contribute to its overall complexity and survival. Complexity in the form of digital sequences is generally perceived as a source of encoded information. The sequence of nucleotides in mRNA determines the digital information content of the molecules—just as the sequence of letters in words determines the information content of words. Digital information in biological systems is primarily understood in terms of the central dogma of molecular biology, an explanation of the flow of genetic information from DNA to RNA to the protein. However, the origin of digital information in prebiotic synthesis is hardly discussed in the literature. In this chapter, we offer a new view on the origin of digital information and the first gene.

Because RNA contains four nucleotide bases analogous to the words' letters, it can function as an information-containing molecule. However, the linear sequence of nucleotides in the early stage of abiotic or nonenzymatic RNA is like a nonsensical word without much meaning. It does not follow any grammar or rules. It could not have specified anything and could not be said to carry much information other than the Watson–Crick base pair rule. The most crucial task in the digital world was to assemble mRNA molecules step by step, creating codons and encoding these molecules with cognate amino acids by tRNAs for protein synthesis.

As mRNA began to encode amino acid information, translation and the genetic code coevolved in the DIS. Side by side, three generations of translation machines in the HIS and three generations of proteins in the AIS evolved hierarchically. Directionality appeared in three types of information systems from the DIS to the HIS to the AIS to complete this cycle (see Fig. 3.1).

Nucleic acids are generally considered sources of encoded digital information, but the origin and evolution of information in molecular systems are not fully understood. The discovery of the genetic code, messenger RNA (mRNA), and the synthetic protein machinery provided a paradigm for thinking about molecular biological processes in computational terms and secured the modern view of biology as an information system [1–4]. Furthermore, Crick's sequence hypothesis and central dogma [5] and Schrödinger's heritable information, a replicable code script [6], established an unquestionable consistency between sequence-based molecular systems of nucleic acids and Darwinian evolution.

Genetic information is essentially digital data plus meaning [7]. The nucleotide sequences of mRNA provide the data, and the meaning is the translation of the data into a functional protein. Translation is transferring information from the language of mRNA to the language of proteins [5–7]. During translation, mRNAs serve as a data storage system, transmitting digital instruction to molecular machines, the ribosomes, which manufacture protein molecules. RNAs are essential for encoding and implementing information. Three kinds of RNA molecules orchestrate translation: the mRNAs encode the recipe for proteins in the digital format and carry this information to the ribosomes via charged tRNA molecules, where the proteins are synthesized. The tRNAs read the message of mRNA and fetch the specific amino acids to the ribosomes during protein synthesis. Finally, the catalytic components, the ribosomal RNAs (rRNAs), translate the nucleotide sequences of mRNAs into amino acid sequences of proteins.

The language of life was born in the peptide/RNA world during the onset of mRNA-dependent protein synthesis. It is a bilingual language in which codon sequences of mRNA are translated into amino acid sequences of proteins. The actual translation process from mRNA into protein language occurs where amino acids and tRNAs are matched and joined. The translators are aaRS enzymes that do this job. They are the only bilingual elements in the translation machine: they can recognize both amino acids and tRNAs' anticodons.

Interestingly, each aaRS recognizes all the tRNAs of the given amino acid. The genetic code is like a translation dictionary that translates 64 codons to their 20 amino acid equivalents, of which 61 represent amino acids and 3 are stop signals. The flow of information from mRNA to proteins began the digital information age.

Translation is the process of decoding the sequence of mRNA molecules to a sequence of amino acids during protein synthesis. Specifically, the 'language' of the nucleotide sequence of an mRNA of the 4 letters (A, U, G, C) is translated into the 20 letters of amino acids alphabets of proteins. Translation is complicated because there is no one-to-one correspondence of the nucleotides of mRNA with the amino acids of the new protein. The four letters of mRNA are read in units of three, called a codon. During translation, an mRNA sequence is read using the genetic code, a set of rules that define how a codon is to be translated into the 20-letter code of amino acids. Availability and interactions of amino acids with noncoding RNAs in the prebiotic hydrothermal vent environment would play critical roles in the emergence of the digital information system step by step.

12.3 Selection of Amino Acids

A total of 70 amino acids were identified in the Murchison meteorite [8, 9]. Miller's experiments created more than 40 different amino acids [10]. Despite the availability of many amino acids in the prebiotic environment, nature has selected 20 amino acids that make up proteins. It is likely that the 20 amino acids needed for protein synthesis were not available initially during prebiotic synthesis; they were selected step by step from a small number of simple abiotic amino acids; these precursors gave rise to more complex amino acids. Wong [11] championed the coevolution theory, and Di Giulio [12] further expanded this model. It proposes that the genetic code is an imprint of the biosynthetic relationships between amino acids. The primitive proteins were created only by those 10 amino acids that were readily obtainable from the prebiotic environment. The remaining 10 amino acids entered the prebiotic system as the code expanded to the universal code.

Ikehara [13, 14] proposed the GADV hypothesis that four primitive amino acids, glycine (G), alanine (A), aspartic acid (D), and valine (V), were available in the prebiotic soup. We suggest that these four amino acids would govern the emergence of four codons (GGC, GCC, GAC, and GUC) by corresponding pre-tRNAs. Because pre-aaRS could recognize both amino acid and its corresponding anticodon of a pre-tRNA molecule, these GADV amino acids would be selected by pre-aaRS enzymes to charge cognate pre-tRNA molecules. These four pre-tRNAs, in turn, would create corresponding codons by base pairing, thus initiating the codon–amino acid mapping (see Sect. 13.6 for a detailed discussion on the role of amino acids in the origin of the genetic code).

12.4 Processing and Assembly of mRNAs: The Emergence of Genes

In the DIS, we see the beginning of creating the coding RNA (mRNA), which was designed and custom-made by a noncoding tRNA. The RNA molecules that initiated protein synthesis belonged to messenger RNA. The mRNA is the most critical molecule for the origin of the DIS, but its birth in the peptide/RNA world remains unknown. In living cells, mRNA is directly transcripted from DNA, where an enzyme (RNA polymerase) converts the gene of a DNA chain into the primary transcript mRNA. However, how did mRNA emerge before the advent of DNA in the peptide/RNA world where there was neither a protein enzyme nor DNA? Here, we propose a model for the origin of the first genes before enzyme and transcription.

There would have been various mechanisms of nonenzymatic polymerization of nucleotides into RNA (e.g., by a condensation reaction in which phosphodiester bonds are formed). However, these abiotic RNAs would have been noncoding RNAs and would have lacked the genetic triplet code for protein synthesis.

A charged tRNA is a key and mobile molecule of the translation machine. It matches with the codons and ferries correct amino acids to the ribosome. It is an adaptor molecule between the codon and amino acid. In our model, it played a key role in the prebiotic synthesis of mRNA. tRNA carries information about its cognate amino acid at one end, and, at the other end, it has an anticodon that can create a codon by base pairing. The dual information system at its two terminals makes tRNA an ideal molecular architect for creating mRNA de novo. Here, we suggest a likely biochemical pathway for the origin of coded mRNAs, which would direct protein synthesis. Aminoacyl tRNAs would create codons in mRNA in two successive steps: (1) by base pairing with anticodons and (2) by encoding the newly formed codons by charged tRNAs, thus transferring information about cognate amino acids. Later, we will discuss how these newly formed codons would be encoded by tRNA containing specific information about amino acids. Codon recognition describes the process of matching codons to the correct amino acids.

12.5 Piecemeal Buildup of Codon Sequences in Pre-mRNA by Charged Pre-tRNA

Codons are the three-letter snippets of mRNA that provide instructions for the specific arrangement of amino acids, which are the building blocks of making proteins. Pre-mRNAs

could be encoded by pre-tRNAs in a step-by-step manner by linking these codons into longer chains. We begin with the pre-tRNAs that might have designed the codons of pre-mRNA molecules for storing digital information. As the pre-tRNA molecules begin to map specific amino acids, they need a separate storage device for the preservation of amino acid information. Pre-tRNAs begin to create custom-made ancestral pre-mRNA strands for the safekeeping of their amino acid information. These pre-mRNA molecules began to map primitive groups of GADV amino acids such as glycine, valine, aspartic acid, and alanine that they tend to encode (Fig. 12.2a).

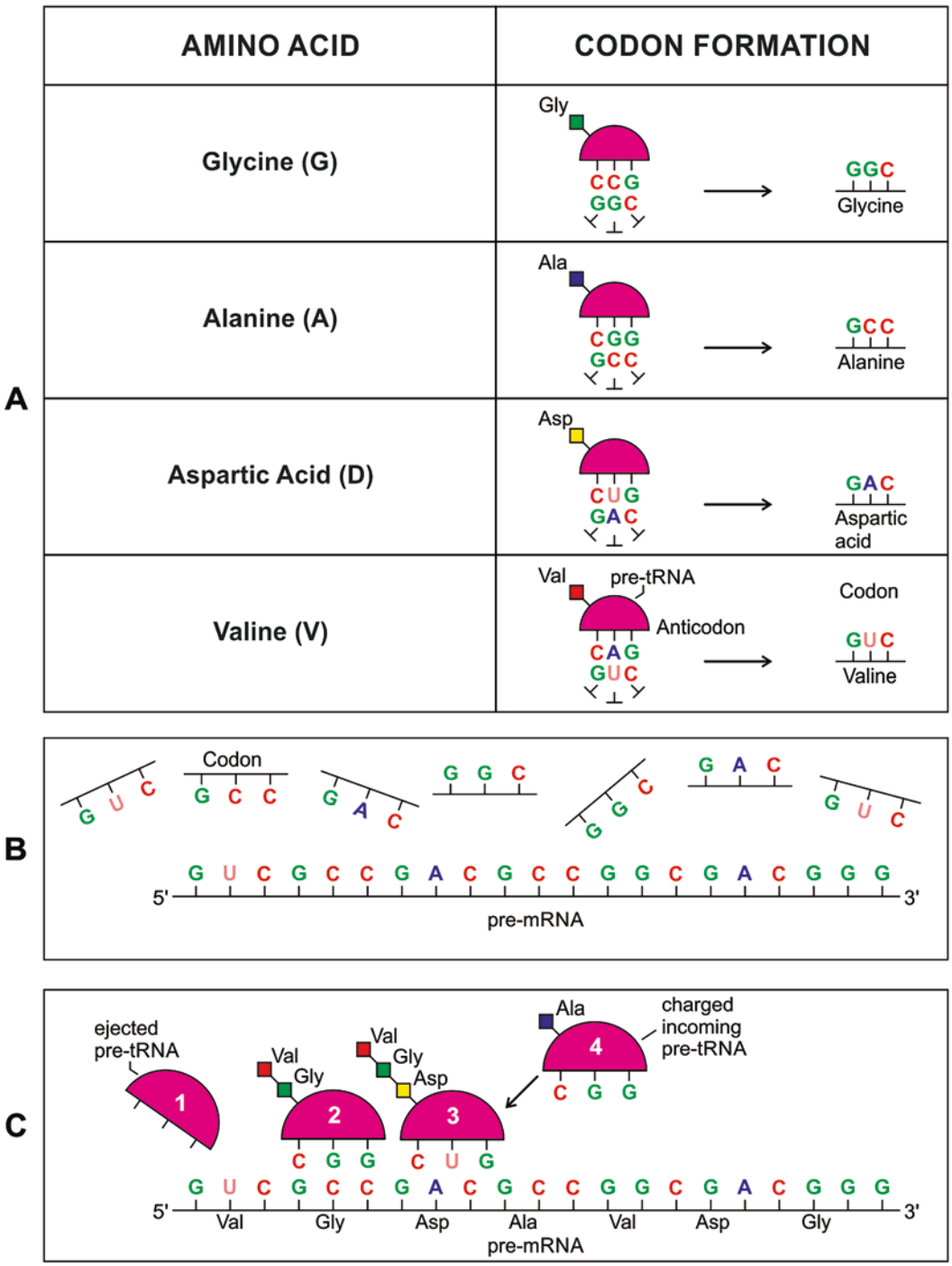


Fig. 12.2 Creation of codons by charged pre-tRNA molecules in a step-by-step manner. (**a**) GADV amino acids govern the origin of codons via pre-tRNAs; an anticodon of a pre-tRNA molecule hybridizing with the corresponding nucleotide available in the prebiotic soup to form a codon strand; each codon develops a memory for a specific amino acid. The four amino acids, glycine (G), alanine (A), aspartic acid (D), and valine (V), were available in the abiotic stage. (**b**) Codons, thus created by charged pre-tRNAs, began to link to form a strand of pre-mRNA with a coding sequence. (**c**) Pre-tRNA and pre-mRNA interactions generated rudimentary translation. In this figure, we offer a specific mapping mechanism between codons and their cognate amino acids that led to rudimentary translation and the genetic code

In our model, pre-tRNA molecules charged with specific amino acids began to select nucleotides from the prebiotic soup via base pairing with their anticodons; these nucleotides were joined to form a codon strand with memory for a specific amino acid, transmitted by the anticodon of a pre-tRNA. The short codon segments, in turn, are polymerized to create a long strand of pre-mRNA for storing digital information about amino acids.

12.6 Random Linking of Codons and Polycodons to the Pre-mRNA Chain

Like nucleotides, codons could be linked to one another in a chain by a condensation reaction to form oligonucleotides and pre-mRNA-like molecules. Olasagasti and Rajamani [15] showed the polymerization of RNA-like molecules experimentally from nonactivated prebiotic mononucleotides and oligonucleotides via condensation reactions. These authors simulated a terrestrial hydrothermal environment that fluctuates between wet and dry cycling during seasonal change. They showed that a mixture of lipids and mononucleotides or oligonucleotides could produce relatively long strands of RNA-like polymers in alternate cycles of dehydration and hydration. It is likely that in a hydrothermal crater vent environment, newly formed codons could link together to pre-mRNA-like molecules by a hydration–dehydration cycle. Clay mineral particles have long been known to have large adsorption capacity and the ability to concentrate and polymerize monomers [16]. The catalytic power of clay minerals can stimulate the polymerization and linking of codons on the floor of the hydrothermal crater basin. The linking of codons could be achieved on the mineral substrate either by the hydration–dehydration cycle at the sloping rim of the crater basin or on the floor of the crater basin (Fig. 6.5). Although 4 codons were formed initially by anticodons of pre-tRNA molecules, 24 codons could be generated by combinations of each of these 4 codons that could create 6 polycodon chains. These 6 polycodons, in turn, could generate 720 polycodons by combinations and link randomly to the form of pre-mRNA molecules (Fig. 12.2b, c). The wet/dry cycle at the hydrothermal crater edge was a natural experimental laboratory for the random assembly of codons and polycodons into a pre-mRNA chain. This pre-mRNA became the binding partner for pre-tRNA, enhancing mutual stability and instant recognition.

As pre-tRNAs created more and more codons by base pairing, the longer chain of pre-mRNAs and mRNAs would be created to store genetic information for the synthesis of long and complex protein molecules. The origin of pre-mRNA and mRNA was the first important step of digital information during life synthesis.

12.7 Encoding the Message

Amino acids do not read their codons. Similarly, codons and amino acids do not recognize each other directly but use molecular intermediaries. The protein and mRNA languages seem unrelated. mRNAs store the amino acid information to assemble proteins using the four-letter alphabets — A, U, C, and G. In contrast, proteins employ 20 different sorts of amino acids. Therefore, how do mRNAs and proteins communicate? How, then, is the message read? In the case of genetic translation, the bilingual tRNA acted as an interpreter and helped translate the message of pre-mRNA, a sequence of codons, to the language of a protein, a sequence of amino acids. Nature has discovered a neat solution to the numerical mismatch by packaging the bases in triplet codons. In all, 4 bases made of 3 nucleotide triplets or codons can be arranged in 64 (4^3) possibilities of codons to code 20 amino acids, and 3 codons are used as stop signals. Consequently, the universal genetic code is redundant or degenerate because more than one codon may code a single amino acid. The genetic code is nonoverlapping, meaning that a single nucleotide cannot share two adjacent codons.

The genetic code was deciphered nearly 60 years ago. Despite decades of effort, it remains largely unknown why certain amino acids are assigned to certain codons. Why do the codons encode specific amino acids during protein synthesis? This is one of the most perplexing questions in molecular biology. There is no direct codon/amino acid interaction during translation, but an anticodon preserves its information of cognate amino acid. Each tRNA molecule has two functional domains: the acceptor domain for bonding and recognizing its cognate amino acid and the second anticodon domain for preserving the memory of its specific amino acid. In an aminoacyl tRNA (or a charged tRNA), its cognate amino acid is chemically bonded, corresponding to its anticodon. A set of aaRS enzymes catalyze the charging of tRNA. No such chemical bond exists between the codon in mRNA and its assigned amino acid. The existence of bilingual adaptors such as charged tRNAs may provide insights into the codon–amino acid association.

Here, we offer a likely mechanism of transfer of information from aa-tRNA to mRNA during the encoding of codons. aa-tRNAs might function as the donor of amino acid memory to the corresponding codons in mRNA. This horizontal transfer of memory of amino acid from charged tRNA to the codons of mRNA is the foundation of the genetic code. mRNA could store long-term memory of amino acids in its codon sequences. In our model, an mRNA is regarded as a ‘memory molecule’ that creates the genetic code step by step during its assembly as a ‘memory bank.’ An mRNA stores the amino acid information in its memory in its codon sequences, whose permanence is established by natural selection.

It is likely that pre-mRNAs are first encoded by charged pre-tRNA molecules and retain the memory of amino acids. Pre-mRNAs would give rise to mRNAs by gene duplication or linking several of their strands. Recycled mRNA molecules after the translation exhibited activity essential for replenishing the nucleotide pool to create a new mRNA chain that would enable a protocell to change its protein expression. Through many generations, natural selection would create encoded mRNAs that were programmed for protein synthesis without the assistance of tRNAs. The memory of codon–amino acid mapping did not fade in the past four billion years during translation in all organisms. It is fascinating how mRNA retained its long-term memory without any sign of decay during the passage of billions of years.

12.8 Encoding Properties of Pre-tRNA and tRNA Molecules

aa-tRNA translation is bilingual according to two different genetic codes. Both pre-tRNAs and tRNAs have two distinct properties for their bilingual proficiency [17, 18]. The accuracy of encoding depends on the precision of the two successive independent matchings. The first matching is manifested in the anticodon of tRNA that created an uncoded codon by the Watson–Crick base pair interaction. This information stage is called 'anticodon–codon mapping' [19]. We used anticodons of tRNA molecules to develop an uncoded sequence of codons, analogous to a blank tape in a tape recorder. Just like how a magnetizable coating is placed on a plastic film to prepare it for storing musical data, the uncoded sequence of codons is prepared to place amino acid information in the second matching. The second matching binds each tRNA with its cognate amino acid in relation to its anticodon, and the charging is catalyzed by a specific aaRS. There are 20 such aaRSs, 1 for each amino acid. Each charged codon then moves its amino acid information to its corresponding codon. We call this second information stage 'codon–amino acid mapping' (Fig. 12.3). The second code is written in the structure of the aaRS and is the likely mechanism of encoding codons with amino acids.

The two-step encoding mechanism of codons by tRNAs left a detectable imprint in stop codons. The three stop codons (UAA, UAG, and UGA), signaling translation termination, may signify that these three blank codons are relics of the first anticodon–codon mapping without an amino acid input.

12.8.1 The Memory Transfer Model

We offer a horizontal memory transfer model from aa-tRNA to mRNA during encoding of mRNA. We use the codon–amino acid mapping scheme to place specific amino acids into their corresponding codons. The blank tape of pre-mRNA and mRNA becomes a recorded tape, mapping each codon onto its specific amino acid. Our model in Fig. 12.3 shows a viable method of encoding codons by tRNAs. We have uncovered a pathway that transferred information from aa-tRNA's temporary memory of a specific amino acid to mRNA in a two-step process. These codons remember amino acids in their codon sequences and store digital information. These encoded codons would develop permanent molecular memory by natural selection. Molecular memory in data storage technologies uses individual molecules (such

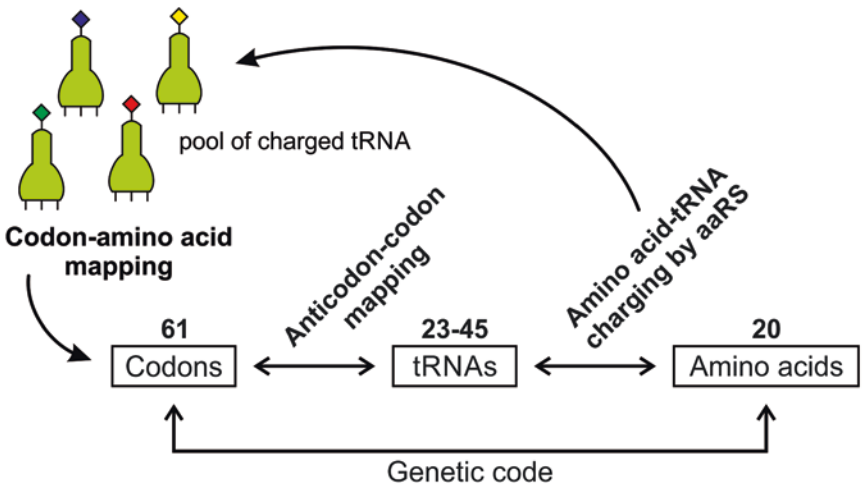


Fig. 12.3 The encoding properties of tRNA. tRNA played two critical roles in creating and encoding codons corresponding to two different genetic codes. First, it created a codon by the Watson–Crick base pair interaction (anticodon–codon mapping). Second, each charged tRNA transferred its amino acid information to the corresponding codon (codon–amino acid mapping). Participation of aaRS in the recognition process is an attractive possibility. aa-tRNA is a transformer, capable of encoding and decoding codons

as zinc porphyrin) for encoding high-density storage data for computers [20]. It is fascinating how aa-tRNA used a similar technology for encoding amino acids data permanently in codon sequences of mRNA four billion years ago. After mRNA translation into a protein chain, mRNA molecules are recycled into nucleotides. These nucleotides, when polymerized to form a new mRNA chain, inherit this memory and become permanently encoded by natural selection. The tRNAs became the matchmakers between codons and their corresponding amino acids. However, codons cannot chemically attach to appropriate amino acids. Moreover, they can store this digital information of amino acids transferred by aa-RNAs in their molecular memory like a tape recorder. The aa-tRNA was a molecular architect that designed and encoded mRNA. If aa-tRNA did not create encoded mRNA, then its bilingual ability to read nucleotide sequences of mRNA and process amino acid information for protein synthesis is difficult to explain. aa-tRNA transferred amino acid knowledge from a source domain to a target domain such as uncoded codons as in a machine learning system. The two-step process of encoding codon maintains the 'specificity' of the coding assignment. The memory transfer model explains why the genetic code assigns similar codons to similar amino acids.

We can gain a new understanding of translation from machine language. In recent years, artificial intelligence programs have reached a surprising level of linguistic fluency, where the learning algorithm is called neurons. The prebiotic information system developed language processing by natural selection four billion years ago, which is still operating in all living organisms with surprising accuracy and fidelity. Like machine language, we propose a progressive 'memory bank' for incremental domain adaptation in codons in three stages of the genetic code. A memory bank stores translation memory in a neural network for machine translation [20]. The mRNA can be viewed as a mobile storage device for storing amino acid information in its codon sequences. Today, encoded mRNA is transcripted and directed from DNA, but in the peptide/RNA world, our model may provide an insight into how the ancestral genes might have evolved before transcription when genetic memory developed in pre-mRNA and mRNA in the linear sequences of codons. The genetic code evolved as a progressive memory bank for codons to remember their amino assignments, and the memory is permanently retained by the natural section. The correspondence between a codon and an amino acid was indirectly realized through amino acid information transfer by tRNA. The memory bank concept may solve the riddle of the specificity of the coding assignments. The transfer of two sets of information from aa-tRNA to mRNA was abandoned and erased from the living digital information system in molecular biology when DNA took over.

12.9 Sequential Evolution of Codon–Amino Acid Assignments

The genetic code defines a mapping between codons and amino acids. The mapping of the codon onto its corresponding amino acid evolved with time. The coevolution theory assumes that the genetic code originated from the metabolic relationships between codons and amino acids [11–14] and will be discussed in detail in the next chapter. The proponents of this theory discovered a beautiful pattern of coevolution between amino acids and the genetic code. However, the missing link in this investigation was the answer to why the pattern was created in the first place. In our view, the pattern of coevolution was created during the encoding process of codons in mRNA by aa-tRNA. Here, we offer a new view of codon assignment.

We follow Ikehara's scheme of the three stages of the genetic code [13, 14] as a guide to simulate the encoding process of codons by aa-tRNA.

- The GNC code: four codons encoded the information of four amino acids
- The SNC code: 16 codons encoded the information of 10 amino acids
- The universal code: 61 codons encoded the information of 20 amino acids

In the primitive stage (the GNC code), the information of four primitive amino acids, glycine (G), alanine (A), aspartic acid (D), and valine (V), was transferred from aa-pre-tRNA to four codons. Codon GGC encoded glycine (G), codon GCC encoded alanine (A), codon GAC encoded aspartic acid (D), and GUC encoded valine (V).

In the next stage (the SNS code), 16 codons, encoding information of 10 amino acids, were transferred from aa-tRNA to 16 codons with the introduction of redundancy so that several codons represented the same single amino acid, but there were no ambiguities. There were no examples of a single codon representing more than one amino acid [12–14]. Among these 16 codons, 12 new codons were embedded with amino acid information by aa-tRNAs. The early genetic code continued to evolve with enhancement of genetic memory.

In the final stage (the universal genetic code), information of 20 amino acids was transferred and distributed by aa-tRNA among 61 codons with a significant component of redundancy. Of these, three codons do not specify amino acids. Three remained uncoded and functioned as stop signals. These stop codons indicate when a polypeptide is complete. The 20 different types of aaRSs can create 20 aa-tRNAs. The coupling of each amino acid by an aaRS to the appropriate tRNA is a pivotal part for the mapping of the genetic code. Thus, aa-tRNAs set the limits of the genetic code. The universal genetic code translates 61 codons into 20 amino acids using fewer than 61 aa-tRNAs. This was because

of the ability of tRNA molecules to 'wobble' at the third base to decode more than one codon because the third base allows less precise (non-Watson–Crick) pairing [21]. The wobble hypothesis explains why multiple codons can code for a single amino acid. From Tables 12.1 and 12.3, it becomes clear that the third base allows wobble or less precise base pairing, when letter codons are converted to numerical codons (1 for U, 2 for C, 3 for A, and 4 for G).

The aa-tRNA has imprinted one line of the genetic dictionary in mRNAs, with all synonyms included. With stable memory sets in 61 codons for 20 amino acids' information and an efficient translation mechanism established by countless natural experiments in the vent environment, the universal genetic code started operating with high fidelity in the peptide/RNA world. In our view, tRNA charging by aaRS and mRNA encoding led to the codon assignments. These two processes are closely linked and took place one after another. The combination of charging and encoding determines the genetic code. The mapping between 64 codons and 20 amino acids created a permanent memory bank, which became the universal genetic code by natural selection.

Genetic memory between a codon and its cognate amino acid is formed by a sequence of repeated interactions in the vent environment, where the process has a starting point, goes through a sequence of steps, and has an end point. Then, the cycle starts all over again. Each process was able to 'remember' its steps and the memory needed to complete the process. These codons 'remembered' which information of amino acids they code and store. This is how genetic memory developed in mRNA that resides in the linear codon sequences. They may be regarded as the memory mapping of the polypeptide structure [3].

The molecular encoding of 61 codons with their 20 cognate amino acids by aa-tRNA molecules was permanently embedded in the codon sequences by innumerable frequencies in the vent environment. The molecular memory is somewhat analogous to machine learning of a bilingual language (e.g., English to German) that lets computers learn to program themselves through experience. In this case, the program is the universal genetic code. In both molecular and machine language translation, information systems are distributed at all levels and operate without any central control [22].

The design of a custom-made pre-mRNA by a pre-tRNA was a watershed event in the origin of translation when a pre-mRNA molecule becomes a storage molecule for genetic information in a separate digital device. Eventually, several strands of pre-mRNA are linked to form a longer strand of pre-mRNA, about 30–80 nucleotides. In the prebiotic soup of the hydrothermal vent environment, varied lengths of pre-mRNA strands by permutation and combination evolved, to store and encode a wide range of amino acid information to synthesize longer protein chains.

12.10 Visualization of Encoding mRNA Molecules: The Origin of Genes

Molecular memory is regarded as a physical mechanism and as a correspondence between an input and the resulting output [3]. In our case, the input is from aa-tRNA, and the output of this event is the codon assignment. We use a novel mechanism of memory transfer of information of amino acid from aa-tRNA to mRNA to show the two-step processes of encoding mRNAs. For ease of visualization, we modify the genetic code in two formats [23]. First, we convert 20 amino acids in 1-letter abbreviations corresponding to their codons in numerical forms. Second, we substitute the nucleotide alphabets of mRNA with numbers as follows: 1 for U, 2 for C, 3 for A, and 4 for G [24] (Table 12.1). We use three additional letters, J, X, and Z, (displayed in bold font), to signify three stop codons, namely, opal, ochre, and amber, respectively.

In Table 12.2, the abbreviation of the universal genetic code table is shown in numerical codons with redundancies. Each matrix cell displays information in both numerical codons and their corresponding amino acids. Because of the numerical distribution of codons in rows and columns, one can easily visualize the distribution of the codons and their redundancies in the matrix cells, which was less evident in a standard genetic code using combinations of four letters. Looking at Table 12.2, we can see those codons beginning with four formed first (the GNC code), followed by codons starting with 2 (the SNS code). Codons with prefixes 1 and 3 were added last to create the universal genetic code.

In a previous publication [23], we had developed a software called CATI (**C**odon–**A**mino Acid **T**ranslator **I**mitator) to translate numerical codons into corresponding one-letter abbreviations of amino acids, and vice versa. Here, we used CATI to map the three stages of amino acid formation (in letters) that corresponds with the code of codons (in numbers) in Table 12.3. The table shows the sequential evolution of amino acid–codon assignments.

In our visualization model, uncoded codons were already created by anticodons of pre-tRNA and tRNA by the Watson–Crick base pair (anticodon–codon mapping) (see Fig. 12.4). This is the first code of pre-tRNA/tRNA. Here, we simulated and visualized the second code of pre-tRNA/tRNA and how these newly generated codons were encoded (codon–amino acid mapping). We showed this codon–amino acid mapping in the three stages of the evolution of the genetic code (Fig. 12.4).

Combining the information provided in the three tables, here, we show the simulation of the encoding processes of mRNA in three stages in tandem with the three stages of the genetic code [13, 14]:

Table 12.1 Decoding table from amino acids (in abbreviated letter) to corresponding codons in numerical forms, showing redundancy. Here, 20 primary amino acids in the genetic code are shown, and their corresponding numerical codons are shown by 23 alphabets. The three letters B, O, and U remain unused

1-Letter Abbreviation	3-Letter Abbreviation	Amino Acid	Numerical Codons
A	Ala	Alanine	421, 422, 423, 424
B	—	—	—
C	Cys	Cysteine	141, 142
D	Asp	Aspartic acid	431, 432
E	Glu	Glutamic acid	433, 434
F	Phe	Phenylalanine	111,112
G	Gly	Glycine	441, 442, 443, 444
H	His	Histidine	231, 232
I	Ile	Isoleucine	311, 312, 313
J	**Stop**	**Opal**	143
K	Lys	Lysine	333, 334
L	Leu	Leucine	113, 114, 211, 212, 213, 214
M	**Met (Start)**	**Methionine**	314
N	Asn	Asparagine	331, 332
O	—	—	—
P	Pro	Proline	221, 222, 223, 224
Q	Gln	Glutamine	233, 234
R	Arg	Arginine	241, 242, 243, 244, 343, 344
S	Ser	Serine	121, 122, 123, 124, 341, 342
T	Thr	Threonine	321, 322, 323, 324
U	—	—	—
V	Val	Valine	411, 412, 413, 414
W	Trp	Tryptophan	144
X	**Stop**	**Ochre**	133
Y	Tyr	Tyrosine	131, 132
Z	**Stop**	**Amber**	134

Stage 1: Visualization of the Encoding Codons of Pre-mRNA by Pre-tRNAs in the GNC Code

In the start cycle of the GNC code, pre-tRNAs created four codons by hybridization: 443 (GGC), 422 (GCC), 432 (GAC), and 412 (GUC) (shown by white circles). In the end cycle, these codons were encoded by charged pre-tRNAs, one at a time, with four cognate amino acids, including G (glycine), A (alanine), D (aspartic acid), and V (valine). This is often called the GADV model based on the availability of primordial amino acids [13, 14].

Stage 2: Visualization of the Encoding Codons of mRNA by tRNAs in the SNS Code

In the start cycle of the SNS code, four codons in blue circles were already encoded in the GNC code. In all, 12 white circles represent new additional codons that remain uncoded. These are 444 (GGG), 424 (GCG), 414 (GUG), 434 (GAG), 234 (CAG), 212 (CUC), 214 (CUG), 222 (CCC), 224 (CCG), 232 (CAC), 242 (CGC), and 244 (CGG). In the end cycle, these 12 codons were encoded by charged tRNAs in the following order of corresponding amino acids: G (glycine), A (alanine), V (valine), E (glutamic acid), Q (glutamine), L (leucine), L (leucine), P (proline), P (proline), H (histidine), R (arginine), and R (arginine). Here, we see the beginning of redundancy that some amino acids are coded by more than one mRNA codon. For example, the amino acid glycine is specified by the codons 443 and 444 and the amino acid alanine by 422 and 424 (see Table 12.3) [13, 14].

Stage 3: Visualization of the Encoding Codons of mRNA by tRNAs in the UG Code

In the start cycle of the universal genetic code, 16 codons in blue circles were encoded by the GNC and the SNS codes. In all, 45 white circles represent new additional codons that remain uncoded. These codons include 441 (GGU), 442 (GGC), 421 (GCU), 423 (GCA), 431 (GAU), 411 (GUU), 413 (GUA), 433 (GAA), 233 (CAA), 113 (UUA), 114 (UUG), 211 (CUU), 213 (CUA), 221 (CCU), 223 (CCA), 231 (CAU), 241 (CGU), 243 (CGA), 343 (AGA), 344 (AGG), 331 (AAU), 332 (AAC), 141 (UGU), 142 (UGC), 311 (AUU), 312 (AUC), 313 (AUA), 333 (AAA), 334 (AAG), 314 (AUG), 111 (UUU), 112 (UUC), 121 (UCU), 122 (UCC), 123 (UCA), 124 (UCG), 341 (AGU), 342

Table 12.2 The universal genetic code showing numerical codons with corresponding amino acids

	1 (U)	2 (C)	3 (A)	4 (G)	
1 (U)	1 1 1, 1 1 2 F (Phe) 1 1 3, 1 1 4 L (Leu)	1 2 1, 1 2 2, 1 2 3, 1 2 4 S (Ser)	1 3 1, 1 3 2 Y (Tyr) 1 3 3, 1 3 4 Z (Stop)	1 4 1, 1 4 2 C (Cys) 1 4 3 J (Stop) 1 4 4 W (Trp)	1 (U) 2 (C) 3 (A) 4 (G)
2 (C)	2 1 1, 2 1 2, 2 1 3, 2 1 4 L (Leu)	2 2 1, 2 2 2, 2 2 3, 2 2 4 P (Pro)	2 3 1, 2 3 2 H (His) 2 3 3, 2 3 4 Q (Gln)	2 4 1, 2 4 2, 2 4 3, 2 4 4 R (Arg)	1 (U) 2 (C) 3 (A) 4 (G)
3 (A)	3 1 1, 3 1 2, 3 1 3 I (Ile) 3 1 4 (Start) M (Met)	3 2 1, 3 2 2, 3 2 3, 3 2 4 T (Thr)	3 3 1, 3 3 2 N (Asn) 3 3 3, 3 3 4 K (Lys)	3 4 1, 3 4 2 S (Ser) 3 4 3, 3 4 4 R (Arg)	1 (U) 2 (C) 3 (A) 4 (G)
4 (G)	4 1 1, 4 1 2, 4 1 3, 4 1 4 V (Val)	4 2 1, 4 2 2, 4 2 3, 4 2 4 A (Ala)	4 3 1, 4 3 2 D (Asp) 4 3 3, 4 3 4 E (Glu)	4 4 1, 4 4 2, 4 4 3, 4 4 4 G (Gly)	1 (U) 2 (C) 3 (A) 4 (G)

(AGC), 321 (ACU), 322 (ACC), 323 (ACA), 324 (ACG), 144 (UGG), 131 (UAU), and 132 (UAC).

In the end cycle, these 45 codons were encoded by charged tRNAs in the following order of cognate amino acids: G (glycine), G (glycine), A (alanine), A (alanine), D (aspartic acid), V (valine), V (valine), E (glutamic acid), Q (glutamine), L (leucine), L (leucine), L (leucine), L (leucine), P (proline), P (proline), H (histidine), R (arginine), R (arginine), R (arginine), R (arginine), N (asparagine), N (asparagine), C (cysteine), C (cysteine), I (isoleucine), I (isoleucine), I (isoleucine), K (lysine), K (lysine), M (methionine), F (phenylalanine), F (phenylalanine), S (serine), S (serine), S (serine), S (serine), S (serine), S (serine), T (threonine), T (threonine), T (threonine), T (threonine), W (tryptophan), Y (tyrosine), and Y (tyrosine).

In the UC code, the redundancy is extreme, where 61 codons code 20 amino acids. For example, the amino acid leucine is specified by six codons UUA, UUG, CUU, CUC, CUA, and CUG.

The visualization model shows how the codons were encoded step by step by charged pre-tRNAs and RNAs in the prebiotic soup, mediated by pre-aaRS and aaRS enzymes. Each codon then developed permanent memory of its cognate amino acid by natural selection. The transfer of amino acid information from an aa-tRNA to a codon provides the 'specificity' of the coding assignments, i.e., which codons code for which amino acids. The model shows the origin of genes step by step before transcription. aa-tRNA was a 'transformer' molecule (a term borrowed from machine language) and used encoder–decoder architecture during creation of the first gene [25]. The encoder consists of encoding layers (like encoding mRNA) that processes the input iteratively one layer after another, whereas the decoder consists of decoding layers that do the same thing to the encoder's output (like decoding mRNA during translation). Ribose sugar is particulalry essential as a building block of RNA, which could have stored the genetic memory of nucleic acid.

Aminoacyl tRNA was a superb molecular architect for designing the first gene. It was also a master programmer for encoding the codon sequences of a gene with the corresponding amino acids to create an elegant genetic code that emerged as a memory bank during codon-amino acid mapping. It used a recursive (self-referencing) computation to generate all possible solutions of 'codon-amino acid mapping' while building

Table 12.3 Codon–amino acid mapping in three stages of the genetic code using CATI software. In the SNS and universal genetic code, the sequence of generating redundancy of codons to amino acids is shown

INPUT OF AMINO ACID INFORMATION BY PRE-tRNA	ENCODING OF CODON SEQUENCE IN OLIGONUCLEOTIDE
GNC CODE	
G	443
A	422
D	432
V	412
INPUT OF AMINO ACID INFORMATION BY tRNA	ENCODING OF CODON SEQUENCE IN SHORT-CHAIN mRNA
SNS CODE	
GG	443444
AA	422424
D	432
VV	412414
E	434
Q	234
LL	212214
PP	222224
H	232
RR	242244
INPUT OF AMINO ACID INFORMATION BY tRNA	ENCODING OF CODON SEQUENCE IN LONG-CHAIN mRNA
UNIVERSAL GENETIC CODE	
GGGG	441442443444
AAAA	421422423424
DD	431432
VVVV	411412413414
EE	433434
QQ	233234
LLLLLL	113114211212213214
PPPP	221222223224
HH	231232
RRRRRR	343344242243244
NN	331332
CC	141142
II	311312313
KK	333334
T	324
FF	111112
SSSSSS	121122123124341342
TTTT	321322323324
W	144
YY	131132

the universal genetic code. The code was embedded directly and permanently into the memory banks of codons. This is why synthetic mRNA or DNA maintains the genetic code during nucleic acid-encoded proteins for research and therapeutic purposes. However, aa-tRNA left no higher-level source code, explanatory comments, documentation, or user manual like modern computer software. Once the universal code was refined by natural selection, all the intermediate steps of creating the code were destroyed to avoid cluttering of raw data, a key to an intelligent information system [26]. By reverse engineering, we can guess how aa-tRNA might have created the digital language of life. With the acquisition of permanent genetic memory, the role of aa-tRNA as an encoder of genes became superfluous. Both mRNA and later DNA became the encoder of genes. aa-tRNA, however, retained the role of decoder for mRNA during translation.

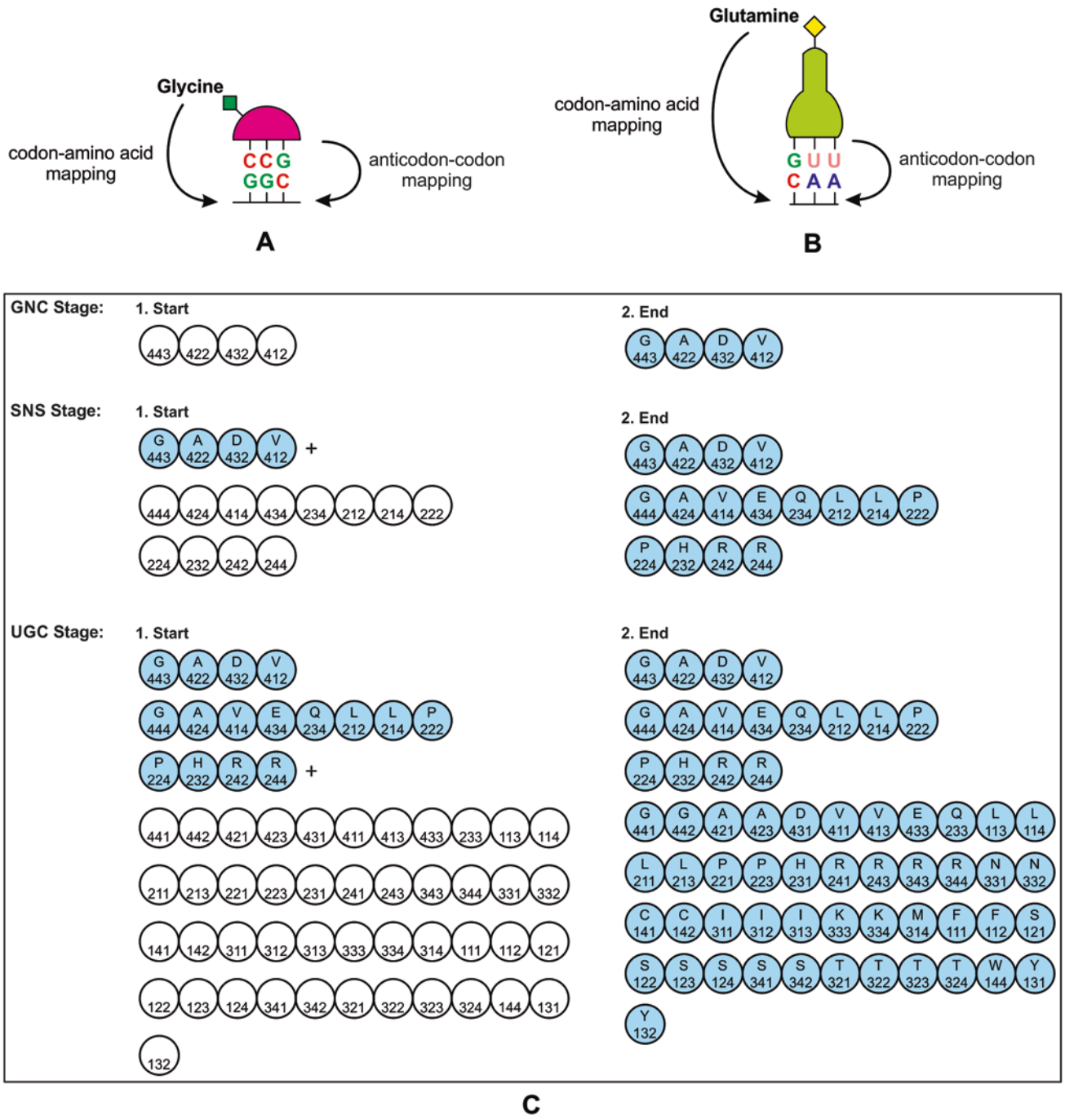


Fig. 12.4 Codon–amino acid mapping and the origin of genes. (**a**) A charged pre-tRNA molecule is an encoding codon of pre-mRNA with glycine as in the GNC code. (**b**) A charged tRNA molecule is an encoding codon of mRNA with glutamine as in the SNS stage. (**c**) Encoding codons by charged pre-tRNA and tRNA molecules in the three stages of the genetic code, controlled by the availability of amino acids in a hydrothermal crater vent environment. In the GNC code, four pre-mRNA codons specify the four amino acids. In the SNS code, 16 mRNA codons code 10 amino acids. In the universal genetic code, 61 mRNA codons designate the 20 amino acids. In the left column of each stage, the white circles represent the uncoded codons, whereas the blue circles represent the encoded codons. In all, 20 aa-tRNA molecules perform the task of encoding codons

12.11 Conclusions

With the emergence of the three components of translation machines such as pre-tRNAs and tRNAs, pre-aaRSs and aaRSs, and, finally, ribosomes, the hybrid information system (HIS) led to the digital information system (DIS). We propose a new 'memory transfer model' for the origin of genes before transcription in the peptide/RNA world. We develop our model of assembling and encoding mRNAs by aa-tRNAs based upon several assumptions and evolutionary principles. The aa-pre-tRNA molecules designed and created custom-made codons for storing amino acid information in three steps: (1) codon formation by aa-pre-tRNAs and aa-tRNAs, (2) linking of codons to pre-mRNAs and mRNAs, and (3) encoding of codons by aa-pre-tRNAs and aa-tRNAs. The DIS is encoded in the linear sequences of nucleotides in pre-mRNA and mRNA. In our model, aa-tRNA has a short-term memory of its cognate amino acid, as long as it remains charged. It aimed to create a separate device, where it can store its information of cognate amino acid. Therefore, it created mRNA that stored long-term memory of amino acids in its codon sequences. During the process of assembly of pre-mRNA and then mRNA, the genetic code emerged as a memory bank during codon–amino acid mapping. mRNA has stored the amino acid information in its memory permanently for the past four billion years in all organisms. Codon–amino acid associations create genetic memory that resides in the codon sequences of mRNA for translation. The encoded mRNA molecules represent the ancestral genes before the advent of DNA. In creating the first gene, the aa-tRNA molecule was a transformer, capable of both encoding and decoding of mRNA.

The central dogma of molecular biology states that 'DNA makes RNA makes proteins.' We are interested in the possibility of intercepting the second step in the peptide/RNA world: 'RNA makes proteins.' Two different RNA molecules, aa-tRNAs and mRNAs, are required in this step for translation. Our new view of the 'prebiotic dogma' of molecular biology states that 'aa-tRNA makes mRNA makes proteins.'

References

1. Walker I (1972) Biological memory. Acta Biotheor 21:203–235
2. Chatterjee S, Yadav S (2022) The coevolution of biomolecules and prebiotic information systems in the origin of life: a visualization model for assembling the first gene. Life 12:834. https://doi.org/10.3390/life12060834
3. Barbieri M (2016) What is information? Philos Trans R Soc 374. https://doi.org/10.1098/rsta.2015.0060
4. Walker IW, Davies PCW (2012) The algorithmic origin of life. J R Soc Interface. https://doi.org/10.1098/rsif.2012.0869
5. Crick FH (1968) The origin of the genetic code. J Mol Biol 38:367–379
6. Schrödinger E (1944) What is life? The physical aspects of living cell. Cambridge University Press, Cambridge
7. Maynard Smith J, Szathmary E (1999) The origins of life. Oxford University Press, New York
8. Bernstein MP, Sandford SA, Allamonda LJ (1999) Life's far-flung raw material. Sci Am 263:42–49
9. Pizzarello S, Cronin JR (2000) Non-racemic amino acids in the Murchison and Murray meteorites. Geochem Cosmochem Acta 64:329–338
10. Bada JL (2013) New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. Chem Soc Rev 42:2186–2196
11. Wong JTF (1975) A coevolution theory of the genetic code. Proc Natl Acad Sci U S A 72:1909–1912
12. Di Giulio M (2008) An extension of the coevolution theory of the origin of the genetic code. Biol Direct 3:37. https://doi.org/10.1186/1745-6150-3-37
13. Ikehara K (2002) Origins of gene, genetic code, protein and life: comprehensive view of life systems from a GNC-SNS primitive code hypothesis. J Biosci 27:165–186
14. Ikehara K (2021) Towards revealing the origin of life: presenting the GADV hypothesis. Springer Nature, Cham
15. Olasagasti F, Rajamani (2019) Lipid-assisted polymerization of nucleotides. Life 9:83. https://doi.org/10.3390/life9040083
16. Kloprogge JT, Hartman H (2021) Clays and the origin of life: the experiments. Life 12:259. https://doi.org/10.3390/life12020259
17. De Duve C (1998) Transfer RNAs: the second genetic code. Nature 333:117–118
18. Mace K, Gillet R (2016) Origins of mRNA: the missing link in the birth of protein synthesis? Nucleic Acids Res 17:8041–8805
19. Roth AC (2012) Decoding properties of tRNA leave a detectable signal in codon usage bias. Bioinformatics 28:340–348. https://doi.org/10.1093/bioinformatics/ibis403
20. Yasseri AA, Syomin D, Malinvoskii VL, Loewe RS, Yu H, Lindsey JS, Zaera F, Bocian DH (2004) Characterization of self-assembled monolayers of porphyrins bearing multiple thio-derivatives rigid-rod tethers. J Am Chem Soc 126:11944–11953
21. Crick FH (1966) Codon-anticodon pairing: the wobble hypothesis. J Mol Biol 19:548–545
22. Navlakha S, Bar-Joseph Z (2011) Distributed information processing in biological and computational systems. Commun ACM 58:94–102
23. Chatterjee S, Yadav S (2019) The origin of prebiotic information system in the peptide/RNA world: a simulation model of the evolution of translation and the genetic code. Life 9:25. https://doi.org/10.3390/life9010025
24. Ashgar N, Mou L, Selby KA, Pantasdo KD, Poupart P, Jiang X (2020) Progressive memory banks for incremental domain adaptation. arXiv:1811.00239
25. Hagiwara M (2021) Real world natural language processing: practical applications with deep learning. Manning Publications, New York
26. Kurzweil R (1999) The age of spiritual machines. Penguin Books, New York

A Code Script for Life

13

The genetic code is not a binary code as in computers, nor an eight-level code as in some telephone systems, but a quaternary code with four symbols. The machine code of the genes is uncannily computerlike.
—Richard Dawkins, 1995

The genetic code defines the mapping between codons and amino acids. It is the universal language of life. We have now reached a point in our hypothetical reconstruction of the age of digital information where the translation machine began to decode mRNA and synthesize proteins using the rules of the genetic code. Nature uses 64 triplet codons to encode the synthesis of mRNA-templated proteins in all domains of life using just 20 amino acids, which produces a degenerate and resilient genetic code. A translation machine is required to decode the nucleotide-based language of mRNA into the language of amino acids. Digital information in the linear codon sequences of mRNA bases is translated into amino acid sequences according to the genetic code to create a protein chain in the analog format.

The origin of the genetic code is a fundamental problem in biology and remains the universal enigma [1]. Although many theories have been proposed in the last 50 years to explain why codons are selectively assigned to specific amino acids, empirical data are extremely rare. The centrality of mRNA, rRNA, tRNA, and RNA-binding peptides in the process of translation strongly suggests that the genetic code emerged in the peptide/RNA world. We will discuss the prevailing theories and provide the historical background of the development of the genetic code. Finally, we will suggest how a new informational paradigm may shed light on the origin of the genetic code.

We have divided the digital information system (DIS) into two parts: the first part deals with the origin of the first gene, which is discussed in Chap. 12. The second part is the origin of the genetic code. Without the origin of mRNA, translation and the genetic code would not occur. In our view, the genetic code evolved as a memory bank during the encoding of mRNA by aa-tRNA molecules step by step, analogous to the memory bank of machine translation [2].

13.1 Cracking the Genetic Code

In the twentieth century, one of the scientific breakthroughs was the discovery and cracking of the genetic code. The discovery of the structure of DNA in 1953 was heralded as groundbreaking. However, a few questions remained: How could the information contained within DNA in the 4-letter alphabet of nucleotides be translated into the 20-letter alphabet of amino acids that built proteins? What was the bilingual language, the genetic code, which made the translation from nucleotides to amino acids possible?

By the early 1950s, researchers found that codons in mRNA carry information to build proteins. At that time, proteins appeared to be constructed from precisely 20 kinds of amino acid building blocks and had a mystical appeal. Why does all life use the same 20 amino acids? The physicist and cosmologist George Gamow coined the term ‘genetic code.’ He postulated that sets of 3 nucleic acid chains must be employed to encode the 20 amino acid living cells used to build proteins. With four bases forming 3 base codons, there are $4^3 = 64$ possible codons. Gamow did not know which of his triplets stood for which amino acid. This was left open and would have to be discovered by a series of novel experiments. By the 1960s, Robert Holley and his team developed techniques to separate tRNA from the cell and established how tRNA incorporates the amino acid alanine to form specific proteins.

The flow of information for protein synthesis was well-understood by 1961: from DNA to mRNA to proteins, but the detailed mechanisms were missing. The next riddle was deciphering the language of DNA, the code. Cracking the code began in 1961 when Marshall Nirenberg and Johann Matthaei at the National Institute of Health, working on codon assignments, announced that a specific sequence of 3 nucleotide bases (codons) determined each of 20 amino acids. They used an in vitro translation system with a synthetic

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_13

polyribonucleotide of a known sequence. They made a special synthetic RNA polyuridylic acid whose only base was uracil (often called the poly-U experiment), a long sequence of nothing but the nucleotide uracil, repeated several times. They obtained a protein consisting of the amino acid phenylalanine, and the protein made was polyphenylalanine, that is, phenylalanine repeating itself over and over again in a long chain. They interpreted that the codon UUU specified the amino acid phenylalanine. Nirenberg went on to decipher the code by showing the correspondence between various codons and individual amino acids. This experiment demonstrated that mRNA transcribes genetic information from DNA, regulating the assembly of amino acids into complex proteins.

After Nirenberg cracked the first 'word' of the genetic code, scientists raced to translate the unique code words for each amino acid, hoping to someday read the entire code of living organisms. This was followed by experiments of earlier Nobel laureate Severo Ochoa's laboratory that demonstrated that the poly-adenine RNA sequence (poly-A) coded for the polypeptide poly-lysine and that the poly-cytosine RNA sequence (poly-C) coded for the polypeptide poly-proline. Thus, the codon AAA specified the amino acid lysine, and the codon CCC specified the amino acid proline. The plot for solving the code mystery was foreshadowed.

Har Gobind Khorana mastered the synthesis of nucleic acids and helped decipher the mechanisms by which RNA codes for the synthesis of proteins. Khorana showed that the repeating nucleotide sequence UCUCUCUCUCUC encodes a strand of amino acids reading serine–leucine–serine–leucine. By 1965, primarily due to Nirenberg and Khorana's work, the genetic code had been completely cracked. It revealed that each codon group encodes a specific amino acid and that the order of the codons determines the order of the amino acids in the resulting protein. Because of their remarkable achievement, Robert Holley with Nirenberg and Khorana shared the Nobel Prize in physiology and medicine in 1968 'for their interpretation of the genetic code and its function in protein synthesis.' The three stop codons were later named and discovered by Richard Epstein and Charles Steinberg, whereas one start codon was found by Marilyn Kozak. Decoding the genetic code has paved the way for tremendous advances in molecular biology and genetics.

13.2 Characteristics of the Genetic Code

A code is a set of rules used to change information from one form to another. For example, the Morse code converts a sequence of dots and dashes into a sequence of letters that can be read as a normal language. The genetic code is often contrasted with the binary code that is used by computers. Two types of codes show significant similarities and differences. Each system has its advantages and limitations. The primary alphabet used in computer code is the binary digit (0,1), or a bit, a contraction for the binary digit. The smallest unit of information or data in a computer is the bit. This source code is translated into computer code by a computer to perform its tasks. Binary information is organized into sets of eight bits or bytes. Each byte is just eight bits, and it is the smallest unit of memory in many computer systems. A byte has one of 2^8 or 256 possible configurations of zeroes and ones. A binary source alphabet could be enlarged by creating ordered pairs, ordered triplets, ordered quadruplets, and so forth; it allows t to form receiving alphabets larger than two.

In the genetic code, these extensions are called codons. On the other hand, the most straightforward unit of mRNA/DNA is the nucleotide, which can have one of four bases (A/T, U, C, and G)— the quaternary 'bit.' The application of the word 'bit' in a quaternary system of mRNA or DNA is a misnomer. Here, we choose a more appropriate name in the genetic code, called 'quit,' or quaternary digit, instead of a bit, the classical bit's genetic version. The quits are a quaternary digits of nucleobases (A/T, U, C, and G). This increased variation suggests that each nucleotide of mRNA can hold twice as much information as each digit of a binary program. The quit creates more algorithmic randomness than a bit, and it is more information-rich. A 'qubit' or quantum bit in quantum computing is the basic unit of quantum information [3].

Both binary and genetic codes contain start and stop signals. Computers start and stop bits for beginning and ending the reading of their messages, respectively. The genetic code, on the other hand, uses one start codon and three stop codons for this purpose. In a binary code, an erroneous bit creates its byte to have a different value, which can cause significant errors. However, the genetic code exhibits greater flexibility, which often accommodates this error, and is more robust.

The universal language of life is the genetic code. It is used by living cells to translate information encoded in the codons of nucleic acids (DNA or mRNA) into proteins. Like a dictionary, it is based on a set of rules. It is a correspondence between codons displayed by mRNAs and amino acids. This defines the rules by which information stored in DNA and transcribed in mRNA sequences is translated into the corresponding amino acid sequences to produce proteins. The genetic code portrays the three-letter words of the mRNA language in a four-letter alphabet of nucleotides to 20 amino acids in the protein language. The numerical mismatch between the 4 bases in mRNA and 20 amino acids in proteins is neatly solved by the genetic code.

There are several properties of the genetic code. Of the 64 possible codons, 61 designate amino acids, indicating that many amino acids are encoded by more than 1 codon. The genetic code is 'specific' (unambiguous), that is, a particular codon always codes for the same amino acid. The code is 'degenerate' or redundant because a single amino acid may be coded by more than one codon (Table 13.1A, B). The

Table 13.1 (A) The universal genetic code table. Each codon is written from 5′ to 3′. The start codon (AUG) is shown in green. The stop codons (UAA, UAG, and UGA) are shown in red. (B) In the genetic code, 20 amino acids are used in protein synthesis, showing corresponding codons in redundancy

First letter (5′)	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter (3′)
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met (Start)	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

A

AMINO ACID	RNA CODON
1. Alanine	GCA, GCC, GCG, GCU
2. Arginine	AGA, AGG, CGA, CGC, CGG, CGU
3. Asparagine	AAC, AAU
4. Aspartic acid	GAC, GAU
5. Cysteine	UGC, UGU
6. Glutamic acid	GAA, GAG
7. Glutamine	CAA, CAG
8. Glycine	GGA, GGC, GGG, GGU
9. Histidine	CAC, CAU
10. Isoleucine	AUA, AUC, AUU
11. Leucine	UUA, UUG, CUA, CUC, CUG, CUU
12.Lysine	AAA, AAG
13. Methionine	AUG (START CODON)
14. Phenylalanine	UUC, UUU
15. Proline	CCA, CCC, CCG, CCU
16. Serine	AGC, AGU, UCA, UCC, UCG, UCU
17. Threonine	ACA, ACC, ACG, ACU
18. Tryptophan	UGG
19. Tyrosine	UAC, UAU
20. Valine	GUA, GUC, GUG, GUU
STOP CODONS	UAA, UAG, UGA

B

more often used amino acids in proteins are specified by a higher number of different codons. No codon goes unused.

The code is also 'comma-free.' The remaining three codons—UAA, UGA, and UAG—do not specify amino acids. Instead, these triplets are 'stop codons' that signal the end of the coding sequence for an mRNA. There are two punctuation marks: an AUG indicates the start and one or more stop codons at the end. The genetic code has no 'ambiguity' but maintains 'redundancy.' For example, codons GAU and GAC both specify aspartic acid (redundancy), and neither of them specifies any other amino acid (ambiguity). The genetic code is 'nonoverlapping,' without overlaps or gaps. During the translation, the codons are read consecutively.

The code is 'ordered.' Multiple codons for a given amino acid and codons for amino acids with similar chemical properties are closely related, usually differing only by a single nucleotide [5, 6]. The sorting of the genetic code is distinctly nonrandom, and neighboring codons are assigned to amino acids with similar physical properties. Hence, the effects of translation errors are minimized concerning reshuffled codes.

The genetic code also provides for the 'punctuation' of genetic information at the level of translation. The AUG is used as a start codon to initiate polypeptide chains. Three codons—UAG, UAA, and UGA—specify polypeptide chain termination (Table 13.1B).

The genetic code is virtually 'universal,' that is, the specificity of the genetic code has been conserved from the very early stages of evolution, with only slight differences in the way the code is translated. Most living organisms follow the same universal genetic code. The most notable exceptions to the code's universality occur in the mitochondria of mammals, yeasts, and several other species. That the code is universal is exceptionally significant, for it suggests it was used by the last universal common ancestor (LUCA) and is robust enough to have survived through four billion years of evolution. Without this inherent stability and robustness, protein production would be haphazard, not tailored, and accurate [4–6].

Mutations are errors caused in codons by changes in nucleotide bases. Changing a single nucleotide base on the mRNA chain can lead to a point mutation. The most common mutations take place in two ways: (1) a base substitution, in which one base is substituted for another, and (2) insertion or deletion, in which a base is either incorrectly inserted or deleted from a codon. Although most mutations that change protein sequences are harmful or neutral, some mutations have benefits. Statistical studies have shown that deleterious mutations and translation error effects are lower

in the universal genetic code than in almost all rearranged codes using the same set of amino acids. The error-minimizing theory gained some credit when it was shown that the universal genetic code is quite robust against mutations [7].

The genetic code is the universal language of life. The code is like a bilingual unidirectional dictionary (because it translates codons into amino acids), not the other way (translation of amino acids to codons is not possible). It is sometimes called the 'Rosetta Stone of life.' In the genetic code, the digital information in mRNA is translated into analog information in proteins [8]. The universal genetic code is shown in its entirety (Table 13.1A, B). It should be observed that the left-hand vertical column in Table 13.1A indicates the first (5′) letter (position) in a codon, the horizontal bar across the top indicates the second letter, and the right-hand vertical column indicates the third (3′) letter. The start and stop codons appear in green and red, respectively. In Table 13.1B, all the 20 amino acids and their corresponding codons are shown.

A reading frame is demarcated by the initial triplet of nucleotides from which translation starts. The codons form 'words' of the message without any commas. This is known as an 'open reading frame.' Every sequence in mRNA or gene is read in its 5′ → 3′ direction in three reading frames. The code has been established by several experimental methods [9, 10].

13.3 The Origin of the Genetic Code

The origin of the genetic code remains controversial, even though the full codon catalog was deciphered more than 50 years ago [1]. Despite diverse opinions and controversies on the origin of the genetic code, several researchers agree on a few basic points [6, 11]:

1. The code developed gradually.
2. Early proteins were made from a small number of amino acids, perhaps as few as four. The present number of 20 was reached progressively by the incorporation of relatively few new amino acids.
3. Despite the small number of initial amino acids, the first code used triplet codons and not the doublets that might have sufficed.
4. The genetic code arose from preexisting amino acid–RNA interactions that enhanced the catalytic activities of ribozymes.
5. Amino acids are excellently arranged in the genetic code table, as amino acids with similar physical and chemical properties are generally arranged in the same column.
6. Amino acids with different physical and chemical properties are generally arranged in the same row.
7. The genetic code is degenerate or redundant because a single amino acid may be coded by more than one codon.

The accuracy of the genetic code translation in protein synthesis depends on two critical steps: correct decoding of mRNAs and exact creation of aminoacyl-tRNAs. aa-tRNAs are assembled by aaRSs, which match specific amino acids to their corresponding tRNAs, as defined by the genetic code. Thus, the genetic code's crucial feature is the attachment of amino acids to tRNA molecules, a step that is carried out by high-energy assignment enzymes, such as aminoacyl-tRNA synthetase (aaRS). We suggest that this attachment first occurred between the ribozyme and amino acid by a bridge peptide, then pre-tRNA and a specific amino acid by pre-aaRS, and, finally, by tRNA and a cognate amino acid by aaRS [2, 3].

The code is not the result of a random assignment. It has a structure. There are many unanswered questions regarding the origin of the code. The numerical discord between 4 nucleotides in mRNA and 20 amino acids in proteins was puzzling because 70 amino acids were available in the prebiotic environment [12]. The code might have evolved at a very early stage in the history of prebiotic synthesis in the peptide/RNA world when 20 amino acids in proteins and 4 nucleotides in mRNA may have been optimized by molecular selection for prebiotic synthesis.

Why is the code universal in all organisms? Two extreme theories have been proposed. Perhaps it is simply a 'frozen accident,' a random choice that locked itself in and remained primarily unchanged once the optimal design was reached [4]. Any change of codon reassignment would trigger a mutation that might be lethal. The alternative is the stereochemical theory, which postulates that the origin of the genetic code must lie in the stereochemical interactions between amino acids and codons [5]. It is likely that all life evolved from the last universal common ancestor (LUCA) that used the universal genetic code. Since then, the universal code has remained unchanged for the last four billion years.

The origin of the genetic code is enigmatic; herein, we propose an evolutionary explanation: the demand for a wide range of protein enzymes over peptides in the prebiotic reactions was the primary selective pressure for the origin of the information-directed protein synthesis. Since amino acids were the building blocks of proteins, the selection of amino acids from the prebiotic environment and the subsequent biosynthetic paths would be crucial to the origin of the genetic code. The role of amino acids in the peptide/RNA world could have been the enhancement of prebiotic reactions.

The origin of the genetic code must address three separate questions. (1) What is the order of inclusion of amino acids into the code? (2) What is the mechanism behind the assignment of amino acids to codons? (3) What primitive genetic

code gave rise to the universal code, and how the primitive code has evolved to the modern genetic code? Here, we discuss two prevailing theories.

1. *The stereochemical theory*: According to this, codon assignments are dictated by the physicochemical relation between cognate anticodons or codons and the corresponding amino acids; perhaps tRNA molecules matched their corresponding amino acids according to their stereochemical affinity [1, 5, 11, 13–20]. This hypothesis proposes that anticodons or codons may directly bind to amino acids.
2. *The coevolution theory*: This suggests that the code structure coevolved with the amino acid biosynthesis pathways [21–24]: the 10 primary or primitive amino acids available in the prebiotic conditions that gave rise to the remaining 10 amino acids derived from the first set. What happened afterward is that some primitive systems acquired the ability to manufacture secondary amino acids and, eventually, primary amino acids.

Here, we discuss these two leading views on the origin of the genetic code, the stereochemical theory and the coevolution theory, in detail. We integrate all the prevailing views on the origin of the genetic code and conclude that the code's organization is the outcome of the coevolution of genes and the genetic code.

13.4 The Stereochemical Theory

The stereochemical theory states that there is a stereochemical basis for the assignment of a given codon to an amino acid. This theory has a long and checkered history. George Gamow, a brilliant physicist, proposed the stereochemical theory together with the very first formulation of the numerical mismatch problem between 4 bases and 20 amino acids, nearly 70 years ago. A direct experimental demonstration of the interaction between amino acids and cognate codons or anticodons was performed by Woese [5]. Progress in aptamer technology has provided a boost to the stereochemical hypothesis.

Because codons and amino acids do not directly recognize each other but instead must interact via chemical intermediaries, there is no apparent reason why a particular codon should match a specific amino acid. Modern translation encodes amino acid sequences without a direct codon/anticodon interaction. Suppose that there was an intrinsic affinity for RNA sequences toward amino acids before the development of the modern genetic code. In that case, there must have been a historical transition in which direct interactions were abandoned. The stereochemical hypothesis provides a glimpse of these ancient ties between RNA sequences and amino acids. It postulates that a physicochemical affinity between amino acids and cognate anticodons or codons is determined by the structure of the code [15–20]. For each amino acid, there is a coding sequence for which it has the highest association. The association between these sequences and amino acids influenced the form and content of the genetic code.

The chemical interactions between RNAs and amino acids are well-known. The negative charges of the nucleotide phosphate attract the positive charges of the basic amino acids, which initiates countless interactions between them. However, these interactions are generalized and can take place between any nucleotide and in any amino acid and in no way account for the 'specificity' of the coding rules. The stereochemical theory posits that in addition to the standard chemical reactions, there are specific affinities between codons and amino acids. In other words, the stereochemical theory may provide a clue for codon assignments [25].

At least two research lines show that some amino acid–(anti) codon pairs have a stereochemical binding. One focuses on binding by evolved aptamers' site; the other focuses on binding by triplet or minihelices to the amino acids. RNA aptamer experiments support the stereochemical theory, in which RNA molecules evolved to bind specific amino acids [18]. The binding site contains (is enriched in) the codon and/or anticodon for the cognate amino acids. This was shown for arginine, histidine, tryptophan, and isoleucine. Such experiments suggest the close association of amino acids with codons.

It is most likely that specificities for some amino acids could derive from stereochemical relations. Thus, the stereochemical hypotheses can account for the codon assignment of Arg, His, Ile, Phe, Tyr, Trp, Gly, Ser, Ala, Cys, and Met. No such interaction was found for Gln and Val, and no research has been conducted for Lys, Asp, Glu, Thr, Asn, and Pro. These amino acids or a subset of them should have been the first to enter the code.

Other experiments indicate that anticodons are selectively concentrated near their corresponding amino acids in the ribosomal structure and that such a concentration is correlated with the universal genetic code [26]. Ribosomal anticodon–amino acid concentration suggests that specific codons were redistributed during the evolution of the code. Thus, anticodon–amino acid interactions shaped the evolution of the genetic code.

Recently, Koonin [28] has proposed that the frozen accident hypothesis of Crick [4] is the default theory of evolution because it does not suggest any specific interactions between amino acids and cognate codons or anticodons. He believed that stereochemical interactions between codons and amino acids might have been important at some early stages of code evolution. He outlined an experimentally testable scenario

for the evolution of the code, in which amino acids are recognized via unique sites in the tertiary structure of proto-tRNAs, rather than by anticodons, i.e., expansion of the code via pro-tRNA duplication.

However, the stereochemical theory has some drawbacks. First, the relationship is difficult to recognize in the present day. If codon/amino acid recognition started this interaction initially, then it could give rise to protein recognition. Second, as demonstrated in experiments, about 60% of amino acids have close to random specificities for the aptamer–amino acid reciprocal reaction [27]. Only arginine recognizes its specific codon. However, arginine is a latecomer and is not even on the list of initial codon formation (Table 13.2). It would be difficult to detect the whole correspondences, especially between a codon/anticodon and an amino acid with a small side chain, which should be used in the most primitive genetic code.

Table 13.2 The evolution of the universal genetic code in three distinct stages. Abbreviations of the four bases: adenine (A), uracil (U), cytosine (C), and guanine (G). (The GNC code represents the following nucleotides: G = G; N = A, U, C, G; and C=C.) Here, the evolution from a GNC code (4 codons) through an SNS code (16 codons) to the universal genetic code (64 codons) is shown. (The SNS code represents the following nucleotides: S denotes G and S; N = A, U, C, G.) A (after [30, 33), B (after [33]); (C) the universal genetic code. Instead of conventional representation, the modern genetic code reflects the order of codon occurrence from GNC to SNS to the modern code (columns G and U are switched to show the sequential code evolution from GNC to SNS to universal)

The GNC primitive genetic code
[4 codons]

	U	C	A	G	
G	Alanine	Alanine	Aspartic acid	Serine or Glyine	C

	U	C	A	G	
G	Valine	Alanine	Aspartic acid	Serine or Glyine	C

The SNS primitive genetic code
[16 codons]

	U	C	A	G	
G	Valine	Alanine	Aspartic acid	Serine	C
			Glutamic acid	Glutamic acid	G
C	Valine	Glutamic acid	Glutamic acid	Glutamic acid	C
					G

(Di Giulio 2008)

	U	C	A	G	
G	Valine	Alanine	Aspartic acid	Glycine	C
			Glutamic acid		G
C	Leucine	Proline	Histidine	Arginine	C
			Glutamine		G

(Ikehara 2002)

The universal genetic code
[64 codons]

Second letter

First letter	U	C	A	G	Third letter
G	Valine GUU GUC GUA GUG	Alanine GCU GCC GCA GCG	Aspartic acid GAU GAC	Glycine GGU GGC GGA GGG	U
			Glutamic acid GAA GAG		C
					A
					G
C	Leucine CUU CUC CUA CUG	Proline CCU CCC CCA CCG	Histidine CAU CAC	Arginine CGU CGC CGA CGG	U
			Glutamine CAA CAG		C
					A
					G
A	Isoleucine AUU AUC AUA	Threonine ACU ACC ACA ACG	Asparagine AAU AAC	Serine AGU AGC	U
	Methionine AUG (start codon)		Lysine AAA AAG	Arginine AGA AGG	C
					A
					G
U	Phenylalanine UUU UUC	Serine UCU UCC UCA UCG	Tyrosine UAU UAC	Cysteine UGU UGC	U
	Leucine UUA UUG		(stop codon) UAA UAG	(stop codon) UGA	C
				Tryptophan UGG	A
					G

Universal code
SNS code
GNC code

13.5 Coevolution Hypothesis

The physiochemical properties of amino acids and their biosynthetic relationships might have played critical roles in the origin of the genetic code. It seems highly likely that the first amino acids were those that were simplest to form by nonbiological chemistry in the prebiotic environment. The coevolution theory suggests that the primitive genetic code used a small number of abiotic simple amino acids. As more complex amino acids were manufactured from these precursors, some codons that encoded a precursor were ceded to its more complex products. Wong [21–24] championed the coevolution theory. This theory proposes that primitive proteins employed only 10 amino acids that were readily available from the prebiotic environment. The remaining 10 amino acids were recruited with the expansion of the code by biosynthesis from the early generation of amino acids. The formation of amino acids by biosynthetic pathways guided the development of the genetic code. The more amino acids could be recruited, the wider selection of proteins could be generated for abiogenesis.

In the coevolution theory, two generations of canonical amino acids were recognized, depending on whether they were available in the prebiotic environment (Phase 1) or were biosynthetically produced (Phase 2) [21–24]. Each phase contains 10 amino acids. Phase 1 amino acids consist of Gly, Ala, Ser, Asp, Glu, Val, Leu, Ile, Pro, and Thr. These primitive amino acids were available in the prebiotic vent environment, delivered by meteorites. The ranks of amino acids in this list strongly correlate with the free energy available in the vent environment for their syntheses: the most thermodynamically efficient is on the top of the list. These 10 amino acids are considered old, and they were represented in the first stage of protein synthesis [24]. They would play important roles in the primitive GNC–SNS code (Fig. 11.2). The crucial point is that only Phase 1 amino acids are produced in laboratory experiments that simulate prebiotic conditions.

Higgs and Pudritz [29] documented that the earliest amino acids in the code are the most abundant in meteorites and prebiotic synthesis experiments. By combining measurements from several meteorites and interplanetary dust particles, and experiments in hydrothermal vents, they concluded that 10 amino acids found in the prebiotic environment could be ranked in decreasing order of frequency as Gly, Ala, Asp, Glu, Val, Ser, Ile, Leu, Pro, and Thr. These amino acids correspond to Phase 1 amino acids of the coevolution theory. The other 10 biological amino acids of Phase 2 are not found in these nonbiological situations.

Biosynthetically sourced amino acids of Phase 2 include Phe, Tyr, Arg, His, Try, Asn, Gln, Lys, Cys, and Met. In Phase 2, the amino acids entered the code employing biosynthesis from the Phase 1 amino acids, resulting in the emergence of tRNA molecules, aminoacyl transferase enzyme, and ribosomes [21–24]. The codons assigned to the Phase 2 amino acids differ very little from the codons of their precursor amino acids. The coevolution theory predicts that codons of precursor product amino acids should be contiguous (separated by a single base change). With the availability of 20 amino acids, the genetic code would be expanded and stabilized in the universal code (see Table 12.1 for the list of 20 canonical amino acids and their 3-letter abbreviations).

Di Giulio [30] modified the early phase of the coevolution theory and concluded that the first amino acids came from biochemical pathways directly originating from sugar degradation. Thus, Ala (from pyruvate), Asp (from oxaloacetate), and Ser or Gly (from phosphoglycerate) could have been the first amino acids to be recruited. In the early GNC code, biosynthetic relationships existed between six sibling amino acids. Ala-Ser, Ser-Gly, Asp-Glu, and Ala-Val played a crucial role in the earliest phases of the genetic code.

Higgs' four-column theory for the origin of the genetic code is based on the possible pathways of amino acids added to the code [31]. A likely early step in code evolution would be a four-column code in which all codons have the same middle base code for the same amino acid. The first four primitive amino acids assigned to the four-column were Gly (G), Ala (A), Asp (D), and Val (V), which were available in the prebiotic environment. The theory assumes that these amino acids are assigned to codons with G at the first position, considering the first code may have only used these codons. The code rapidly developed into a four-column code, where all codons in the same column coded for the same amino acid. Afterward, amino acids were sequentially added to the code by reassigning a block of codons with similar physicochemical properties and preserving the error-minimizing property of the code. The selection of amino acids was the driving force behind the code evolution. In this regard, the four-column theory is highly explicit on the fitness advantage of the inclusion of novel amino acids.

Several hypotheses predict the order of inclusion of amino acids in the genetic code. These orders tend to be consensual and are considered structurally simple amino acids, including both early and late complex ones. There is a causal connection between the gradual evolution of amino acids and the code structure (see Sect. 12.9 in the previous chapter). Ikehara [32, 33] selected amino acids in the code evolution based on protein structure formation related to the origin of the genetic code. He argues that the genetic code mediates a codon sequence of the gene with an amino acid sequence of a protein. Therefore, it is important to understand what amino acids were used in the first genetic code for protein synthesis. He proposed the GNC–SNS primitive genetic code hypothesis, which gave rise to the universal genetic

code. We have used Ikehara's tripartite code evolution in the previous chapter during the encoding of the gene.

Ikehara split the early phase (Phase 1) of amino acids into two stages corresponding to the evolution of the genetic code: the first stage (the GADV hypothesis) comprises four primitive amino acids—Gly (G), Ala (A), Asp (D), and Val (V)—for the GNC code. The second stage consists of 10 cumulative amino acids (Phase 1) that gave rise to the SNS code. The incremental third stage (Phase 1 + Phase 2) includes all 20 amino acids that correlate with the universal genetic code. The universal code evolved to have 10 more complex amino acids, which are also more diverse in physiochemical terms than Phase 1 amino acids. This order of inclusion of amino acids suggests constraints toward increasing the codon assignments and the genetic code's versatility for diverse types of specialized proteins (Table 13.2).

13.6 Integrated Coevolution Theory

The genetic code, which associates codons with amino acids, has a unique feature that reflects its function. This is the code's gradual acquisition of amino acids, both prebiotically and biotically. The step-by-step addition of amino acids allows genes to encode a greater repertoire of proteins. We suggest that further affirmation of the role of coevolution can be achieved by including the role of genes in the origin and evolution of the genetic code. Storage and transfer of information of amino acids in mRNA is the key requirement for the origin of the genetic code. The genetic code defines the rule set to decode this amino acid information stored in the codon sequences of early genes.

Here, we integrate the coevolution theory in the light of the digital information paradigm. The coevolution theory provided likely pathways of the genetic code's evolution but was vague about its origin. In the previous chapter, we proposed a new 'memory transfer model' from aa-tRNA to mRNA during its encoding process [2]. The order of availability of amino acids in the prebiotic environment created the sequence of the amino acid recruitment of aaRS for charging tRNA. The correct recognition of individual amino acids is a crucial determinant for aaRS for charging tRNA. Each aa-tRNA not only created a codon by base pairing but transferred its amino acid information to the newly created codon. This is how the 'specificity' of the codon–amino acid assignment was maintained in the codon sequences of mRNA (Fig. 12.4). There is a code within a code. The code of aa-tRNA (Fig. 12.3) has the same logical structure as in the genetic code and played crucial role in its origin. Thus, the appearance of amino acids via new biochemical pathways was strongly coupled to their integration into the assembly of the gene by aa-tRNA. With the genetic memory embedded in the codon sequences in mRNA, a mapping between each codon and its cognate amino acid was developed. The genetic code evolved as a memory bank during the encoding of mRNAs by aa-tRNA molecules step by step in three stages, analogous to the progressive memory bank in machine language [34]. As more and more codons 'remembered' their cognate amino acids, the memory was permanently etched in the codon sequences by natural selection. The mapping between codons and amino acids developed the genetic code in their 'memory bank.' Thus, aa-tRNA is a hidden code, a palimpsest within the genetic code, and the availability of amino acids in the hydrothermal vent and the subsequent biosynthesis created the order of aa-tRNA for gene making.

The origin of the genetic code is difficult to explain without discussing the origin of genes, translation machines, and protein synthesis. Although the coevolution theory provides the pattern of the evolution of the genetic code, mRNAs or early genes offer the structural basis for the origin of the genetic code. We suggest that the code's organization is the outcome of two distinct coevolution phenomena. These are (1) the coevolution of the translation machine and the genetic code and (2) the coevolution of genes and the genetic code. We follow Ikehara's tripartite evolution of the genetic code, namely, GNC–SNS–UC, in our discussion [32, 33].

13.6.1 Coevolution of the Translation Machine and the Genetic Code

Translation involves decoding an mRNA and using its digital information to build a polypeptide or a chain of amino acids. The actual step of translation from mRNA or genes into protein language occurs when anticodons of aa-tRNAs are matched with codons of encoded mRNAs. The translators are charged tRNAs. If the creation and encoding of codons are molecular rehearsals, then decoding of codon translation is the actual acting of aa-tRNA on the stage. Codon recognition became easy by aa-tRNA in translation. The two processes of tRNA codes (Fig. 12.3) replay during mRNA translation in two successive independent matchings. First, the tRNA anticodon detects the codons by base pairing; it reads the message according to the genetic code. Second, each charged tRNA ferries cognate amino acids to the ribosome and is linked to form a polypeptide chain as the mRNA passes through and is 'read' by the ribosome.

Woese [5] suggested that code assignments and the translation mechanism evolved together. Here, we expand his idea and document the coevolution of the translation machine and the genetic code. In Chap. 11, we suggested how noncoding RNAs created the critical components of a translation machine. During the evolution of the genetic

code, the DIS, the hybrid information system (HIS), and the analog information system (AIS) are so intermingled and interdependent that they worked in tandem in precision: The DIS for the enrichment of the genomic content of mRNA, the HIS for refining the translation machine, and the AIS for the manufactured product of translation and protein synthesis. The unidirectional flow of information developed from the DIS to the HIS to the AIS (see Fig. 3.1).

We suggest that three stages of the development of the genetic code [32, 33] coevolved with three stages of translation machines in the prebiotic world [3]. These are as follows:

- The GNC code by the pre-tRNA/pre-mRNA machine
- The SNS code by tRNA/aaRS machine
- The UG code by the tRNA/aaRS/ribosome machine

The pre-tRNA/pre-aaRS machine started the GNC code, the transitional SNS code progressed by the tRNA/aaRS machine, and the universal genetic code blossomed by the tRNA/aaRS/ribosome machine. The ribosome is the latest addition to the translation machine. It automatically joins an amino acid to a growing polypeptide chain. It is a small, dense particle comprising a large and a small subunit, each made of RNAs and protein components. Here, we identify three stages of the coevolution of the translation machine and the genetic code (Fig. 13.1).

Translation machines are an extremely complicated hierarchy of complex macromolecules that are symbiotically related to one another. Yet, the whole interactive molecular

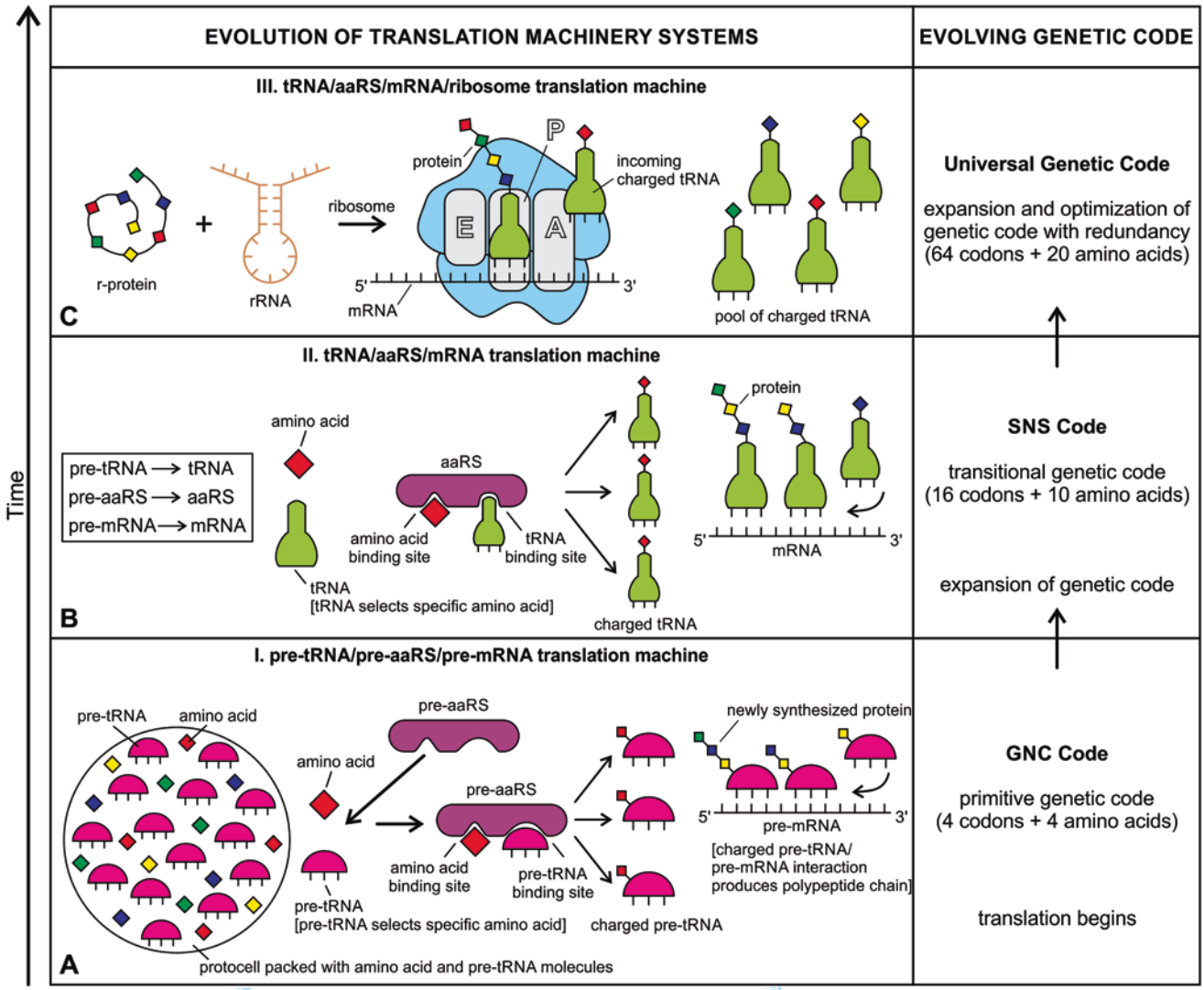


Fig. 13.1 The coevolution of translation machines and the genetic code in three stages: (**a**) Decoding of a pre-mRNA molecule by a pre-tRNA/pre-aaRS translation machine when the GNC code evolved; (**b**) decoding of a short-chain mRNA molecule by a tRNA/aaRS translation machine when the SNS code appeared; and, finally, (**c**) decoding of a long-chain mRNA by a tRNA/aaRS/ribosome machine when the universal code evolved. With the improvement of the translation machine, the information density of mRNA also increased. (Modified from Chatterjee and Yadav [3])

system functions with remarkable precision. Once the translation machinery complex for protein synthesis is installed step by step, digital information enters the system via symbiotic interactions of mRNA with tRNA, aaRS, and the ribosome. Side by side, the genetic code expanded more and more. In Chap. 11, we have summarized how such complex translation machinery would evolve step by step into today's protein-synthesizing machinery, starting from the ribozyme (Fig. 11.3).

13.6.2 Coevolution of Genes and the Genetic Code

The seed of the genetic code is implicit in the emergence of the first gene by aa-tRNA, as discussed in the previous section. Here, we summarize some of the salient points about the origin of the genetic code considering the new information paradigm. In our view, the genetic code evolved as a memory bank during the encoding of mRNAs by aa-tRNA molecules step by step, analogous to the memory bank of machine translation. The genetic code is the memory bank of codon–amino acid mapping. The mRNA-directed protein synthesis was the culmination of the digital information system (DIS) in the peptide/RNA world, where the HIS translation machine read the message of the encoded mRNA of DIS to produce protein in the AIS format. The rules of the genetic code guided protein synthesis. Here, we discuss the pattern and process of the evolution of the genetic code.

We show the coevolution of genes with genetic codes in three stages during protein synthesis:

- Pre-mRNA and the GNC code
- Short-chain mRNA and the SNS code
- Long-chain mRNA and the UG code

Coevolution of Pre-mRNA and the GNC Code
In the early DIS stage, the primitive translation machine of pre-tRNA/pre-aaRS began to decode a short pre-mRNA strand consisting of four codons (GGC, GCC, GAC, and GUC), encoding four GADV amino acids (glycine, alanine, aspartic acid, and valine) and creating a short chain of the biosynthetic protein. This is the first simple polypeptide chain made by the DIS (Figs. 13.1a and 13.2a). In this stage, the primitive GNC code [32, 33] appeared.

Coevolution of Short-Chain mRNA and the SNS Code
In the next stage of the DIS, pre-mRNA evolved into mRNA through gene duplication or linking several strands of pre-mRNA to increase storage capacity. mRNA recruited 16 codons (GGC, GGG, GCC, GCG, GAC, GAG, GUC, CUC, GUG, CCC, CCG, CAC, CAG, CGC, and CGG) or some combination of these codons to code 10 amino acids (glycine, alanine, aspartic acid, valine, glutamic acid, leucine, proline, histidine, glutamine, and arginine). Decoding of mRNA was performed by a tRNA/aaRS translation machine. These modifications of the DIS and HIS gave rise to the SNS transitional code [50, 51] (Figs. 13.1b and 13.2b). The superior information-bearing qualities of mRNA, the excellent catalytic potential of aaRS, and the improved adaptor capacities of tRNA emerged with the gradual expansion of the genetic code. At this stage, tRNAs selected and recruited six additional amino acids. The expanded SNS code was refined through the symbiotic interactions of the tRNA/aaRS complex. The evolution of tRNA and aaRS considerably improved the translation system from the GNC to the SNS stage, but the code remains only moderately robust and susceptible to errors because of the limited redundancy.

Coevolution of Long-Chain mRNA and the UC Code
In the final stage of the DIS, longer strands of mRNA evolved to store the recipe of complex proteins. Most likely, the first mRNA genes were short, no longer than 70–100 nucleotides (the modern genes run several thousand nucleotides), with the corresponding proteins [27]. The translation machine was enhanced with the incorporation of the ribosome, an enormous hybrid of rRNAs and r-proteins. The ribosome improved the efficiency of translation, leading to the universal genetic code with its 20 amino acids and 64 codons (Figs. 13.1c and 13.2c; Table 13.2). In the UC code, 10 precursor amino acids (such as glycine, alanine, aspartic acid, valine, glutamic acid, leucine, proline, histidine, arginine, and glutamic acid) and 10 more derived amino acids (such as isoleucine, methionine, threonine, asparagine, lysine, serine, phenylalanine, tyrosine, cysteine, and tryptophan) were used for encoding long-chain mRNA (Fig. 13.2c). These 20 amino acids became the alphabets for proteins. The long-chain mRNA was decoded by the tRNA/aaRS/ribosome translation machine, giving rise to a complex protein chain.

13.6.3 Flow of Information from Nucleic Acids to Proteins

The climax of the peptide/RNA world was the evolution of mRNA-directed protein synthesis via a translation machine. This was the first breakthrough in life to begin. Life depends on cycles of information between nucleic acids and proteins. In the DIS, we see the progressive improvement of mRNA with more information content and the emergence of the genetic code. In the HIS, there was a refinement of the translation machine. In the AIS, 20 amino acids were generated for protein synthesis. Three prebiotic information systems, the DIS, HIS, and AIS, worked in harmony for this remarkable achievement of making proteins (Table 13.3).

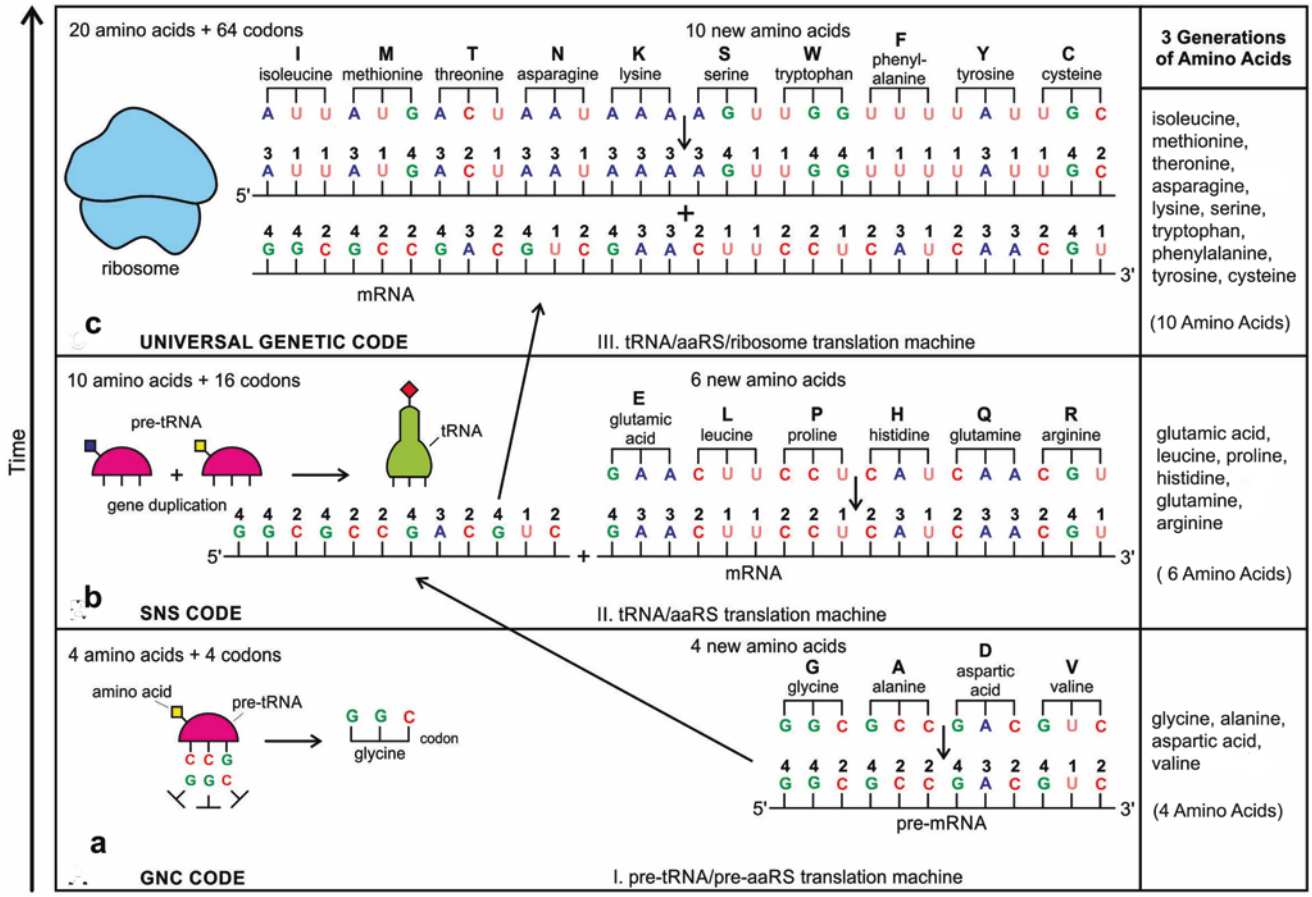


Fig. 13.2 Three stages of the evolution of mRNA, translation machines, and the genetic code. (**a**) Decoding of a pre-mRNA by a pre-tRNA/pre-aaRS machine, resulting in the primitive GNC code. (**b**) Decoding of a short-chain mRNA by a tRNA/aaRS machine in the transitional SNS code. (**c**) Decoding of a long-chain mRNA by a tRNA/aaRS/ribosome machine in the universal genetic code. The right column of the diagram shows the recruitment of amino acids during the evolution of the genetic code

Table 13.3 Three stages of the DIS, HIS, and AIS during the evolution of the genetic code. In the GNC code, the pre-mRNA was decoded by a pre-tRNA/pre-aaRS translation machine, creating a polypeptide chain. In the SNS code, a short-chain mRNA was decoded by a tRNA/aaRS machine, producing a short-chain protein. In the universal genetic code, a long-chain mRNA was decoded by a tRNA/aaRS/ribosome machine, thus manufacturing a long-chain protein

Time

3 Generations of mRNA (DIS)	3 Generations of Genetic Code (DIS)	3 Generations of Translation Machine (HIS)	3 Generations of Protein Chain (AIS)
Long-chain mRNA	Universal Genetic Code	tRNA/aaRS/ribosome	Long-chain protein
Short-chain mRNA	SNS Code	tRNA/aaRS	Short-chain protein
pre-mRNA	GNC Code	pre-tRNA/pre-aaRS	Polypeptide chain

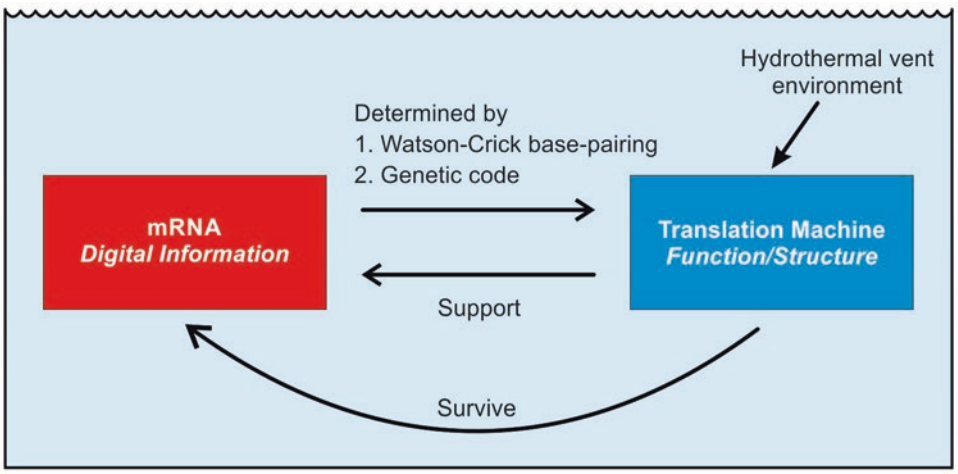


Fig. 13.3 Darwinian evolution began in the peptide/RNA world, an interplay between digital information and its supporting structure, such as a translation machine. The supporting structure is coupled to the information carrier by rules, such as RNA base pairing and the genetic code. The supporting structure is nourished by the chemicals and energy from the hydrothermal vent environment and provides the information carrier with positive feedback

13.7 The Onset of Darwinian Evolution

Darwinian selection started very early in the origin of life and probably played a major role in abiogenesis. However, there is no consensus on when and how the Darwinian evolution emerged during abiogenesis. We suggested [3] that the emergence of translation machines was the beginning of the Darwinian evolution, the reciprocity between information and its supporting structure. This view has been recently elaborated by Kunnev [35], who advocated the proposal that Darwinian evolution began when information carrier molecules, such as mRNA, and its corresponding supporting structure, such as a translation machine, began to interact, where information determines its supporting structure. The structural component is coupled to the information component through rules such as the Watson–Crick base pair and the genetic code that maps changes in the information carrier onto changes in the supporting structure. The structural component is nourished by environmental chemicals and energy that provide information carriers with a positive feedback (Fig. 13.3). Kunnev proclaims that abiogenesis reached 'a point of no return' at this digital stage to enhance life synthesis with the onset of Darwinian evolution.

13.8 Encoding and Decoding of Digital Information

The coevolution theory is reflected in the tripartite evolution of the genetic code, namely, the GNC, SNS, and universal code. It offers a valuable guide to exploring the hidden player in the origin of the genetic code, namely, the gene. Encoding of mRNA created the genetic code as a memory bank (origin) but decoding of mRNA processed the genetic code for protein synthesis (translation). Protein synthesis is the expected action of the decoding mRNA, an attribute of the digital information system. Thus, there are three components in the genetic code: origin, evolution, and translation.

The use of informational terms is widespread in molecular biology. Here, we apply several informational terms such as 'coder,' 'transmitter,' 'receiver,' 'decoder,' and 'information channel' to show the encoding and decoding of mRNA during protein synthesis [36]. A digital communication system built upon the information from mRNA has emerged from our study, where aa-tRNA was the 'transformer' for both encoding and decoding mRNA, as shown in Fig. 13.4 [37]. Initially, during the buildup of the digital information system (DIS), the 'encoder of information' was aa-pre-tRNA/aa-tRNA that created codons by base pairing and transferred specific amino acid information to the codon. Once aa-tRNA fully encoded all the codons in mRNA, there was a division of labor. mRNA became the 'encoder' of information on amino acids in the codon sequences by natural selection. Aminoacyl tRNA continued to perform the role of the decoder of mRNA in collaboration with the ribosome.

The sugar-phosphate (S-P) backbone in mRNA emerges as the key structural framework in transmitting digital messages from mRNA to the ribosome during translation (Fig. 13.4a). The S-P backbone consists of alternating 5-carbon sugar ribose and phosphate groups tightly linked to form a chain and defines the directionality of the molecule. Attached to each sugar is one of four bases – adenine, guanine, uracil, and cytosine. These nitrogenous bases are arranged as appendages all along this backbone and can be assembled end-to-end into linear molecules. The backbone has a 5′ to 3′ polarity where one chain end differs greatly from the other. In an mRNA strand, one end has an unlinked 5′ carbon on the pentose ribose, while the other end has an unlinked 3′ carbon. When the molecule is synthesized, the sugar-phosphate bond extends or grows, which coincides with the flow of the digital information system.

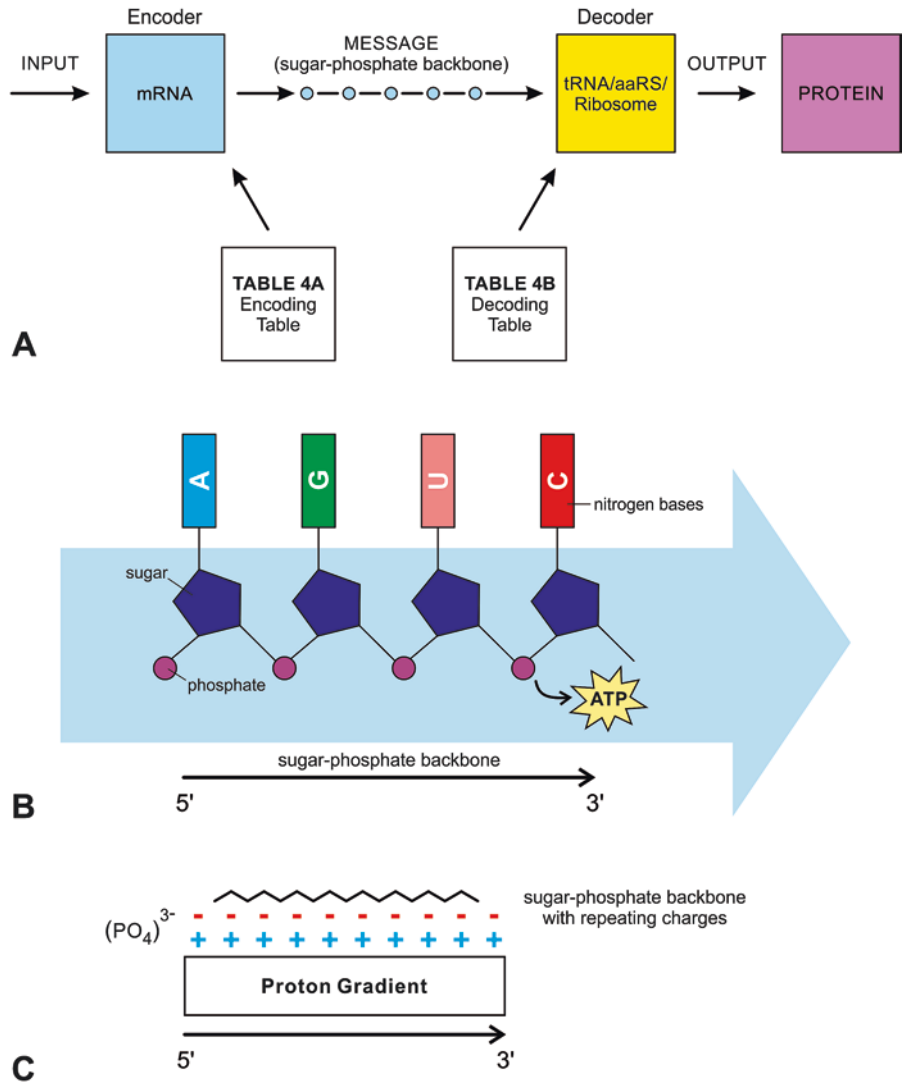


Fig. 13.4 Encoding and decoding of digital information. (**a**) mRNA played critical roles in the origin of the genetic code and translation, where the aa-RNA molecule acted as a transformer to encode and decode mRNA. In a digital information transmission system, aa-tRNA functions as the encoder of mRNA as the input and aa-tRNA in conjunction with the ribosome as a decoder of information creates a protein as the output. The input of encoding of mRNA came from aa-tRNA when the codon assignment was permanently embedded in the codon sequences of mRNA. Once aa-tRNA fully encoded mRNA codons, mRNA became the encoder of amino acids in its codon sequences by natural selection. The digital message from mRNA travels through the sugar-phosphate backbone in a DC electric field. (**b**) adding two phosphate groups with terminal polar phosphate creates ATP that helps to link nucleotides during polymerization. The ribosome decoded the mRNA message to generate protein chain as the output. The sugar-phosphate backbone with repeating charges of phosphate ion, creates a proton gradient similar to connecting rechargeable mini batteries in series together to generate an electrical cable line. (**c**) a sugar-phosphate backbone with repeating charges of phosphate ion, creates a proton gradient similar to connecting rechargeable mini batteries in series together to create an electrical cable, which becomes active like a live wire during translation. The ribosome becomes an electrodynamic machine in contact with the S-P backbone and decodes the message of mRNA to create a protein chain

Phosphates are principal reservoirs of biochemical energy (such as ATP, creatine phosphate, and phosphoenolpyruvate). Operating in water, the repeating negative charge carried by the backbone phosphates in mRNA and DNA is important for many reasons, solubility in water being one [38]. The sugar-phosphate backbone is ionic and negatively charged with H^+ to balance neutrality. It is hydrophilic, allowing strong interactions with water to store and transfer energy, and protects nucleotides that carry genetic information. The ionic phosphate groups of the backbone represent the primary hydration sites for the surrounding water molecules, creating an electric potential between the exterior and interior of the backbone. Polymerization of nucleotides is an endergonic reaction; the energy comes from the phosphate.

So, when mRNA is synthesized, an ATP (adenosine triphosphate), the primary instant energy source released through hydrolysis of its terminal phosphate group, is used to link the mRNA together in a chain (Fig. 13.4b).

Here I hypothesize another critical role of the S-P backbone, an electrochemical strand, for digital communication channel during translation (Fig. 13.4a). Electric interactions with the water molecule and the counterion in the hydrothermal vent environment determine the S-P backbone's structure and dynamics; they provide the energy needed for many essential functions during translation, including the orchestration of translation processes and signal transduction from the environment to the gene via peptide channel of the protocell (see Fig. 22b). All the charged phosphate ions are stacked on one side of an aqueous environment to create an electric potential on the phosphate side of the backbone. One side has a different charge than the other, creating a proton gradient (Fig. 13.4c). That's essentially how a battery works. Each phosphate ion has a voltage, like a tiny battery. The S-P backbone is similar to connecting rechargeable batteries in series together to create an electrical cable, DC electric field, and increase the overall voltage, where *start* and *stop* codons act like its two terminals. The longer the backbone (as in DNA), the greater the voltage.

During the resting state, the S-P channel is inactive. Initiation of translation begins when the protocell receives an environmental signal to activate a specific mRNA strand for making a protein. The ribosome organizes the meeting of mRNA and aa-tRNA molecules to translate the mRNA code into a sequence of amino acids. Ribosome functional sites (RNA-binding sites, decoding center, peptidyl transferase center, peptide exit tunnel) continually exchange information during the various translation steps. The ribosome must synchronize extremely complex movements, such as the rachet-like motion between small and large subunits. In addition, the overall ribosomal dynamics are modulated by the three tRNA site occupancies and their aminoacylation status. Many of the ribosome's complex functional properties go far beyond a mechanical machine's capability. The question is: Who controls the ribosome choreography during protein synthesis?

Ribosomes are among the largest and most dynamic molecular motors. There is growing evidence that the structure and dynamics of the ribosome are more complex during the translation process than it has been assumed for a long time. The physical mechanism of these interactions could be better understood in light of new information on the ribosome. Because most parts of the ribosome are composed of negatively charged rRNAs, its electrostatics should play a fundamental role in protein synthesis [39]. Similarly, r-protein networks in the ribosome play a critical role analogous to nervous-like circuits at a nanoscale [40]. Thus, r-proteins may collectively integrate information taken from distinct sites of the ribosome to synchronize ribosome movements and the correct tRNA recognition and tRNA translocation. But the ribosome needs an external power source – the action of electrochemical input to activate this complex repertoire. Here we propose a causal physical link between the S-P backbone and the ribosome during translation, which generates an electric field to synchronize the ribosome's movements and enable information transfer and processing.

In our model, the S-P backbone becomes active like a live wire during translation in contact with the ribosome, performing like an *information channel* to transmit digital message from mRNA to the ribosome. The ribosome is converted from an electrostatic to electrodynamic machine during contact with the S-P backbone. The small ribosomal unit binds to the mRNA binding site and rachets along the S-P backbone codon by codon in the 5´–> 3´ direction, incorporating each new amino acid and translocating from one codon to the next. The DC electric field generated along the S-P backbone might have synchronized the concerted movement of the ribosome with mRNA and aminoacyl tRNA, all working together like an orchestra during translation (Fig. 13.4c). The ribosome moves along the S-P backbone until it encounters the stop codon when it falls off the mRNA molecule and releases the manufactured protein molecule. In our model, start and stop codons on the S-P backbone act like 'on-and-off' switches during translation. The signal of digital information from mRNA travels through the sugar-phosphate backbone because of the influx and efflux of protons (Fig. 13.4c). The unidirectionality and electrical conduction of the S-P backbone explain for the first time how the digital message can be transmitted from mRNA to the ribosome during protein synthesis.

13.9 Conclusions

The genetic code is the correspondence between codons of mRNAs and amino acids in proteins. Translation of protein-coding mRNAs follows the rules set by the genetic code's codon–amino acid assignments. The codons, in turn, are read by the aa-tRNA anticodons according to the base pairing rules. The crucial information role of aa-tRNA synthetases is to attach amino acids to tRNAs bearing a correct anticodon. The discovery of the structure of DNA in 1953 prompted researchers to unlock the mystery of the genetic code. Scientists like Nirenberg, Ochoa, Khorana, and others finally unlocked the mystery of the genetic code.

The genetic code is a set of rules on how 61 codons code for 20 amino acids and 3 codons are used as stop signals. The code is universal for all organisms, from simple bacteria and eukaryotes to plants, fungi, animals, and humans. The code developed gradually. The origin of the genetic code is still a mystery. There are currently two major theories on the origin and evolution of the genetic code: the stereochemical theory

and the coevolution theory. The stereochemical theory suggests that codon assignments for specific amino acids are determined by a physiochemical interaction between amino acids and cognate codons or anticodons. The coevolution theory asserts that the code structure coevolved with amino acid biosynthesis pathways. The stereochemical interactions between codons and amino acids are possible in some early stages of evolution, but, after decades of research, there is still no real evidence in favor of the stereochemical theory.

The coevolution theory was proposed by J.T.F. Wong and is based on the order of inclusion of amino acids in the code. It is likely that 20 proteinogenic amino acids were not initially available in the prebiotic environment. The theory asserts two generations of canonical amino acids, depending on whether they were readily available in the prebiotic environment (Phase 1) or biosynthetically produced (Phase 2). Each phase contains 10 amino acids. Primordial proteins consisted of only those amino acids that were readily obtainable from the prebiotic environment. The remaining amino acids entered the system as the code expanded along the pathways of amino acid biosynthesis. The genetic code is an imprint of the biosynthetic relationships between amino acids.

Kenji Ikehara subdivided Phase 1 amino acids into stages corresponding to the evolution of the genetic code. The 4 most abundant GADV amino acids created the GNC code, and 10 primitive amino acids gave rise to the SNS code. Finally, 20 amino acids were recruited in the universal genetic code. The code of aa-tRNA is a hidden code within the genetic code and played a crucial role in the origin and structure of the genetic code.

Along with amino acids, genes and genetic codes should have coevolved. We have integrated the coevolution theory by incorporating genes, translation machines, and protein synthesis. We identified two distinct coevolution phenomena using Ikehara's tripartite genetic code evolution. These are (1) the coevolution of the translation machine and the genetic code and (2) the coevolution of genes and the genetic code.

The emergence of translation machines began the Darwinian evolution, an interplay between information and its supporting structure. The genetic code developed in three stages that are coincident with the refinement of translation machines: the GNC code was developed by the pre-tRNA/pre-aaRS machine, the SNS code by the tRNA/aaRS machine, and, finally, the universal genetic code by the tRNA/aaRS/ribosome machine. The ribosome is the latest addition to the translation machine. It is a dense hybrid molecule comprised of a large and a small subunit, each made of RNAs and protein components. It automatically joins an amino acid to a growing polypeptide chain.

The role of the sugar-phosphate backbone in nucleic acid as an electrochemical strand during translation has never been explored. The S-P backbone is analogous to connecting batteries in a series together to generate DC electric field. During translation, the S-P backbone of mRNA might have created a channel or electrical cable that facilitated the transmission of genetic information from mRNA to the ribosome to create a protein chain. The ribosome becomes an efficient electrodynamic machine during translation while gliding along the S-P backbone.

The coevolution theory explains the evolution of the genetic code into three stages, namely, GNC, SNS, and universal code. Still, the origin and processing of the genetic code are linked to the gene. In the previous chapter, we have discussed how encoding mRNA by aa-tRNA created the genetic code as a memory bank. Here, we show three stages of decoding of mRNA by aa-tRNA to process or translate the genetic code during protein synthesis. The integrated coevolution theory may explain the origin, evolution, and translation of the genetic code.

References

1. Koonin EV, Novozhilov AS (2008) Origin and evolution of the genetic code: the universal enigma. Life 61:99–111
2. Chatterjee S, Yadav S (2022) The coevolution of biomolecules and prebiotic information systems in the origin of life: a visualization model for assembling the first gene. Life 12:834. https://doi.org/10.3390/life12060834
3. Chatterjee S, Yadav S (2019) The origin of prebiotic information system in the peptide/RNA world: a simulation model of the evolution of translation and the genetic code. Life 9:25. https://doi.org/10.3390/life9010025
4. Crick FH (1968) The origin of the genetic code. Jour Mol Biol 38:367–379
5. Woese C (1973) Evolution of the genetic code. Naturwissens 60:447–459
6. Osawa S (1995) Evolution of the genetic code. Oxford University Press, Oxford
7. Ardell DH (1998) On error minimizing in a sequential origin of the genetic code. J Morph Evol 47:1–13
8. Koonin EV (2015) Why the central dogma: on the nature of the great biological exclusion principle. Biol Dir 10:52. https://doi.org/10.1186/s13062-015-0084-3
9. Walker SI, Davies PCW (2012) The algorithmic origins of life. J R Soc Interface 10:20120869. https://doi.org/10.1098/rsif.2012.0869
10. Rosen R (1988) Complexity and information. J Comput Appl Math 22:211–218
11. Maynard Smith J, Szathmary E (1999) The origins of life. Oxford University Press, New York
12. Pizzarello JR, Cronin JR (2000) Non-racemic amino acids in the Murchison and Murray meteorites. Geochem Cosmochem Acta 64:329–338
13. Wolf YI, Koonin EV (2007) On the origin of the translation system and the genetic code world by means of natural selection, exaptation, and subfunctionalization. Biol Dir 2:14. https://doi.org/10.1186/1745-6150-2-14
14. Svoboda P, Cara A (2006) Hairpin RNA: a secondary structure of primary importance. Cell Mol Life Sci 63:901–908
15. Woese CR (2002) On the evolution of cells. Proc Natl Acad Sci U S A 99:8742–8747

16. Dunhill P (1966) Triplet nucleotide-amino acid pairing: a stereochemical division between protein and non-protein amino acids. Nature 210:25–26
17. Yarus M, Widmann JJ, Knight R (2009) RNA-amino acid binding: a stereochemical era for the genetic code. J Mol Evol 69:406–429
18. Yarus M (2000) RNA-ligand chemistry: a testable source of genetic code. RNA 6:475–484
19. Yarus M (2021) Evolution of the standard genetic code. Jour Mol Biol 89:19–44
20. Freeland SJ, Knight RD, Landweber LF, Hurst LD (2000) Early fixation of an optimal genetic code. Mol Biol Evol 17:511–518
21. Wong JTF (1981) Coevolution of genetic code and amino acid biosynthesis. Trends Biochem Sci 6:33–36
22. Wong JTF (2014) Emergence of life: from functional RNA selection to natural selection and beyond. Front Biosci 19:1117–1150
23. Wong JTF, Ng SK, Mat WK, Hu T, Xue H (2016) Coevolution theory of the genetic code at age forty: pathway to translation and synthetic life. Life 6:12. https://doi.org/10.3390/life6010012
24. Wong JTF (1975) A co-evolution theory of the genetic code. Proc Natl Acad Sci U S A 72:1909–1912
25. Barbieri M (2015) Evolution of the genetic code: the ribosome-oriented model. Biol Theor 10:301–310
26. Rodin AS, Szathmary E, Rodin SN (2011) On the origin of genetic code and tRNA before translation. Biol Dir 6:14
27. De Duve C (2005) Singularities: landmarks on the pathways of life. Cambridge University Press, New York
28. Koonin EV (2017) Frozen accident pushing 50: stereochemistry, expansion, and chance in the evolution of the genetic code. Life 7:22. https://doi.org/10.3390/life/70200022
29. Higgs PG, Pudritz RE (2009) A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. Astrobiology 9:483–490
30. Di Giulio M (2008) An extension of the coevolution theory of the origin of the genetic code. Biol Dir 3:37. https://doi.org/10.1186/1745-6150-3-37
31. Higgs PG (2009) A four-column theory for the origin of the genetic code: tracing the evolutionary pathways that gave rise to an optimized code. Biol Dir 4:16. https://doi.org/10.1186/1745-6150.4-16
32. Ikehara K (2002) Origins of gene, genetic code, protein and life: comprehensive view of life systems from a GNC-SNS primitive code hypothesis. J Biosci 27:165–186
33. Ikehara K (2021) Towards revealing the origins of life: processing the GADV hypothesis. Springer Nature, Cham
34. Ashgar N, Mou L, Selby KA, Pantasdo KD, Poupart P, Jiang X (2020) Progressive memory banks for incremental domain adaptation. arXiv:1811.00239
35. Kunnev D (2020) Origin of life: the point of no return. Life 10:25. https://doi.org/10.3390/life10110269
36. Maynard Smith J (2000) The concept of information in biology. Phil Sci 67(2):177–194
37. Hagiwara M (2021) Real world natural language processing: practical applications with deep learning. Manning Publications, New York
38. Benner SA (2017). Detecting Darwinism from molecules in Enceladus plumes, Jupiter's moons, and other planetary lagoons. Astrobiology 17(9). https://doi.org/10.1089/ast.2016,1611
39. Wang J, Karki C, Li L (2020). Electrostatics of prokaryotic ribosome and its biological information. Biophys J, https://doi.org/10.1016/j.bpj.2020.01.014
40. Timsit Y, Bennequin D (2019). Nervous-like circuits in the ribosome facts, hypotheses, and perspectives. Int J Mol Sci, 20, https://doi.org/10.3390/ijms200122911

The Advent of Proteins

14

Proteins hold the key to the whole subject of the molecular basis of biological reactions.
—Linus Pauling, 1949

The origin of mRNA-directed proteins in the prebiotic world was a watershed event in abiogenesis. Proteins mediate most functions of modern cells. They are the nanobots of a cell, controlling its many aspects, such as metabolism, transportation, communication, etc. With their astonishing versatility, protein enzymes catalyzed crucial biochemical reactions within protocells into more complex biomolecules in diverse metabolic pathways, whereas structural proteins provided strength and permeability to cell membranes.

The first proteins were most likely short, about 25 amino acids long [1]. Protein molecules of this short length displayed enzyme-like activities. In contrast, many modern-day proteins contain several hundred amino acids. Lengthening probably did not take place gradually base per base, or amino acid per amino acid, but, preferably in a modular manner, by the combination of entire segments of an mRNA. These long molecules of proteins arose by the gradual lengthening of the nucleotide sequences of mRNA or genes that encode the recipe of a specific protein (see Table 13.3).

Proteins are the primary functional biomolecules of life. Once formed, they perform many functions during abiogenesis, including catalyzing metabolic reactions and reinforcing cell membranes. An overwhelming number of efficient proteins were present in the biochemical synthesis that occurred in the protein/RNA world. Once different proteins were synthesized, various mRNAs, ribosomes, and proteins began to accumulate in the cytoplasm, which would be readily available when needed.

The newly synthesized protein enzymes helped catalyze and mediate these critical molecular evolutions, favored by strong selective forces.

Three significant events followed in succession after the availability of template-directed proteins but with considerable overlap. These first affected the efficiency of the translation machinery, then, the resilience of the coding system, and, finally, the quality of the synthesized proteins.

Digital information paved the way for the second wave of the analog information system by manufacturing custom-made proteins according to the specificity of the codon sequences of mRNA. With the emergence of proteins, the first cycle of prebiotic systems ended, which began with an analog information system (AIS) (see Fig. 3.1). The second cycle of prebiotic information systems commenced with the analog information system in the protein/RNA world.

14.1 Prebiotic Protein Synthesis

mRNA-directed protein synthesis is a remarkable repertoire of molecular choreography. Proteins have a versatile chemical structure that allows for the construction of widely different molecular machines using the same basic set of 20 different amino acids, each with a different size and chemical character. In proteins, the amino acids are linked by peptide bonds, the outcome of dehydrating condensation of the carboxyl group (COOH) of one with the amino group (NH_2) of another. By using the different alphabets of amino acids, cells can build a wide variety of proteins. Protein synthesis needs the choreography of dozens of various enzymes. Twenty tRNA molecules, each with its specific synthetase enzyme, are dedicated to 20 amino acids. Modern protein synthesis proceeds with the participation of 20 amino acids, tRNAs, mRNAs, ribosomes, various enzymes, including aminoacyl-tRNA synthetases, ribozymes, peptidyl transferase, a considerable number of protein factors, ATP, GTP, and others [2]. The role of electrochemical sugar-phosphate backbone in the racheting movement of the ribosome and its orchestration with aa-tRNA during translation is discussed in previous chapter (Sect. 13.8). Protein synthesis requires more than 120 species of RNAs and proteins [3] that might have evolved gradually. These biomolecules are related to, encapsulated by, and interacted with each other in complex ways, like an autopoietic machine. They function with

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_14

astounding precision in a kind of molecular choreography in living cells. Perhaps, a similar but more primitive molecular choreography gave birth to mRNA-directed proteins and the universal genetic code about four billion years ago.

Translation takes place in ribosomes (Fig. 14.1). The pathway of protein synthesis translates the codons on mRNA into amino acids that constitute proteins. In the first step of translation, the initial portion of the mRNA binds to a rRNA molecule that is interwoven in the ribosome. The mRNA is positioned so that only a codon segment is exposed at a time to the polypeptide-making site. The mRNA is translated from its 5′-end to its 3′-end, thus producing a protein synthesized from its amino-terminal end to its carboxyl-terminal end [3].

Each gene in the mRNA encodes the recipe for one molecule of protein for multiple functions of the cell, including generating energy, catalyzing a reaction, or any other necessary work. In the ribosome, three stages and three operational sites are involved in the protein production line, and all work in harmony. During the initiation stage, a small ribosomal subunit links onto the 'start end' of an mRNA strand. Aminoacyl-tRNA also enters the A-site of the ribosome. The production of the protein has now been initiated. The second stage, elongation, consists of joining amino acids to the growing protein chain, according to the sequence that was specified by the message. The incorporation of each amino acid occurs by the same mechanism. In the termination stage, the ribosome reaches the end of the mRNA strand, a terminal, or the 'end of the protein code' message. This termination registers the end of production of the specific protein coded by this strand of mRNA (Fig. 14.1). The polypeptide chains produced may be modified by posttranslational modification.

The lengthening of genes and the refinement of translation machines gradually produced new kinds of proteins with new functions. With time, protocells progressively acquired new proteins, especially enzymes, for enhancing catalytic reactions. Proto-metabolism gradually evolved to metabolism. All the analog information were present in the metabolic network. Many cytoplasmic, water-soluble proteins are enzymatic and catalyze chemical reactions involved in metabolism and reproduction. Many of the critical enzymes began to accumulate in the cytoplasm.

14.2 Protein Structure and Function

Proteins mediate most functions of modern cells. During translation, proteins were assembled from amino acids using information encoded in mRNAs. The ensuing peptide elongation was catalyzed by rRNA in the ribosome. Each protein has a unique sequence of amino acids, usually several hundred of them. Once the polypeptide chain synthesis is complete, the polypeptide chain folds to adopt a specific structure that permits the protein to carry out its functions.

Protein molecules control most of a cell's functions: breaking down nutrients, assembling cellular components, and copying DNA, among others. They truly occupied a central position in the organization of the first cell. They acted as enzymes that permitted only a few of the many possible reactions among cellular components to take place. They formed channels in plasma membranes, allowing specific substances to enter and leave while excluding others.

In living cells, a thousand genes make thousands of kinds of proteins of different genes. Each one performs a vital task inside the body, and, collectively, they determine the characteristics of organisms. Protein molecules owe their properties to their three-dimensional shapes, which are determined by the amino acid sequences of their constituent chains. These properties, in turn, determine how a protein functions biologically: whether it will bind specific organic molecules and catalyze their reactions or form a regular structure such as a helix and act as a building material.

Proteins are the primary functional biomolecules of life. Once formed, they perform a vast array of functions during biogenesis, including catalyzing metabolic reactions and reinforcing cell membranes. An overwhelming number of efficient proteins were present in the biochemical synthesis that occurred in the protein/RNA world. With their astonishing versatility, protein enzymes would have taken the role of ribozymes as catalyst and replicase to assist in genetic copying and metabolism. Different kinds of enzymes were in high demand in the protein/RNA world. An enzyme must necessarily have a specific substrate

Fig. 14.1 (continued) tRNA. The tRNA-binding sites are designated E for the exit, P for the peptidyl-tRNA, and A-sites for the aminoacyl-tRNA. The binding site for mRNA is in a small subunit. Translation occurs in a four-step cycle (**c**–**f**) repeatedly during protein synthesis. (**c**) In step 1, an aminoacyl-tRNA, with an appropriate anticodon, enters the vacant A-site on the ribosome, where it hybridizes with a codon. (**d**) In step 2, the carboxyl end of the protein chain is uncoupled from the tRNA at the P-site, then joined by a peptide bond to the free amino group of the amino acid linked to the tRNA at the A-site. This reaction is catalyzed by an enzymatic site in the large subunit, called peptidyl transferase (PT). (**e**) In step 3, a shift in the large subunit (shown by the arrow) relative to the small subunit in the 3′-direction moves the two tRNAs into the E- and P-sites of the large unit and ejects the empty tRNA from the E-site. (**f**) In step 4, the small subunit moves exactly three nucleotides along the mRNA molecule, bringing it back to its original position relative to the large subunit. This movement resets the ribosome with an empty A-site so that the next aminoacyl-tRNA molecule can bind. The cycle repeats when the incoming aminoacyl-tRNA binds to the codon of the A-site. (**g**) The life cycle of the ribosome during its translation [3]

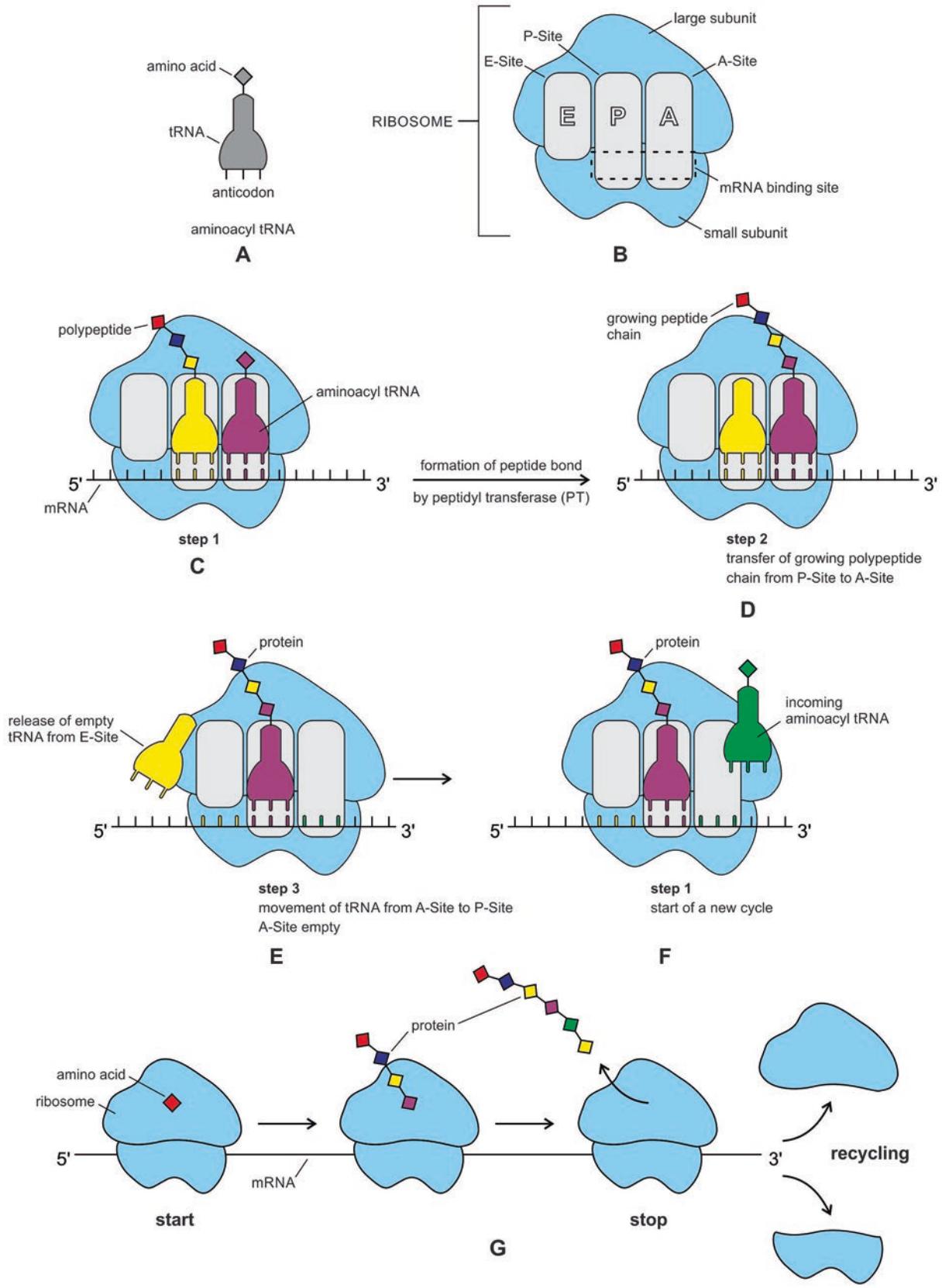


Fig. 14.1 An mRNA is decoded in the ribosome that provides the substrate for controlling the interaction between the mRNA and aminoacyl-tRNA. (**a**) An aminoacyl-tRNA, with an appropriate anticodon. (**b**) Each ribosome has a binding site for mRNA and three binding sites for

available in its environment on which to act. An enzyme must also have an outlet for the products it forms. These substrates and outlets must have been provided by the primitive metabolism in the vent environment that supported the protocells at the time [1]. These newly formed enzymes carried out hundreds of chemical reactions in protocells. Structural proteins, on the other hand, provided structure and support to cell membranes.

14.2.1 Protein Folding

A protein is a long chain of, on average, 300 building blocks of amino acids linked together to form its primary structure. Every protein carries a unique sequence of amino acids. The variation in the amino acid sequence to make proteins is limitless. A typical bacterium makes 50,000 different types of proteins, each with a different function. Protein chains are flexible because the groups on either side of each peptide bond can rotate about their single bonds. The ordering of 20 different types of amino acids determines how the protein chain will fold upon itself. The chain of amino acids can curl, fold, pucker, and zag in three-dimensional space. By folding and coiling into a specific three-dimensional shape, proteins can perform their biological functions. The specific relationship between a protein and other molecules is based on the configuration of the protein. The analog information and the memory contained in the amino acid sequence is sufficient to determine how the chain folds to form the three-dimensional structure of globular proteins. For many proteins, this process occurs simultaneously, but in many cases, special proteins called chaperonins are used to facilitate the folding of the molecules. Proteins fold to their active native state when they emerge from the ribosome. Each polypeptide is carefully folded, creating unique sections that are tailored for their specific job. If a protein unfolds or is denatured, then it will no longer function. A protein's structure, or conformation, is relatively complex and is organized into four categories: primary, secondary, tertiary, and quaternary.

The primary structure is a string of amino acids in a polypeptide chain (Fig. 14.2a). The gene determines the sequence of amino acids that builds the protein. Even changing just one amino acid by a mutation in a protein's sequence can affect the protein's overall structure and function. Although the polypeptide backbone is chemically regular, it contains flexible links; many different shapes are possible in principle. Protein chains in water twist and fold, finding an optimal form to shelter hydrophobic amino acids inside and display charged amino acids on the surface.

The next level of protein folding is called a secondary structure when two types of patterns emerge. Some protein chains coil up into 'alpha-helices,' whereas other regions fold into zigzag patterns called 'beta-sheets,' which resemble the folds of a paper fan. The alpha-helices and beta-sheets form due to interactions between the components of a protein's peptide-bonded backbone. These two structures can interact to create more complex tertiary structures (Fig. 14.2b).

The secondary structure then folds to form a three-dimensional tertiary structure. The tertiary structure of polypeptides results from interactions among R-groups or between R-groups and peptide bonds. It is created by bonds and other interactions that cause proteins to fold in a precise manner. The alpha-helices and beta-sheets can be amphiphilic, which helps form the protein's tertiary structure. Here, the folding occurs so that the hydrophilic sides face an aqueous environment surrounding the protein, and the hydrophobic sides face the protein's hydrophobic core.

The tertiary structure may give way to the formation of a quaternary structure in some proteins. The first three levels of a protein structure are characteristic of a polypeptide chain. However, many proteins contain multiple polypeptides chains, also known as subunits, which interact to form a single structure. The combination of polypeptides as subunits gives proteins a quaternary structure (Fig. 14.2b).

Like nucleic acids, proteins are long molecular chains; however, they are built of 20 different amino acids and therefore contain more information, but the information system of proteins has not been explored.

14.2.2 Enzymes

Enzymes catalyze and significantly speed up chemical reactions that take place within cells. Other than ribozymes, most enzymes are proteins. Without enzymes, the reactions within cells would not happen fast enough to keep the cells alive. An enzyme catalyzes and regulates the rate at which chemical reactions proceed without itself being altered in the process. Enzymes have two fundamental properties like other catalysts. First, they accelerate the chemical reactions without themselves being consumed or permanently altered by the reaction. Second, they increase reaction rates without changing the chemical equilibrium between reactants and products. In general, enzyme names end with the suffix 'ase.'

A large fraction of cytoplasmic, water-soluble enzymes mediate metabolism and reproduction. These enzymes contain a well-packed hydrophobic core that provides structural stability, a predominantly hydrophilic exterior to maintain water solubility, and a catalytic site. Enzymes are proteins that provide a 3D surface for catalysis. Many enzymes are composed of conventional secondary structures such as alpha-helices and beta-sheets, connected by turns and loops.

Enzyme molecules are specific because of their shape and chemical properties; each one catalyzes a particular chemical reaction. The correct folding of an enzyme creates a pocket called an active site that fits only certain

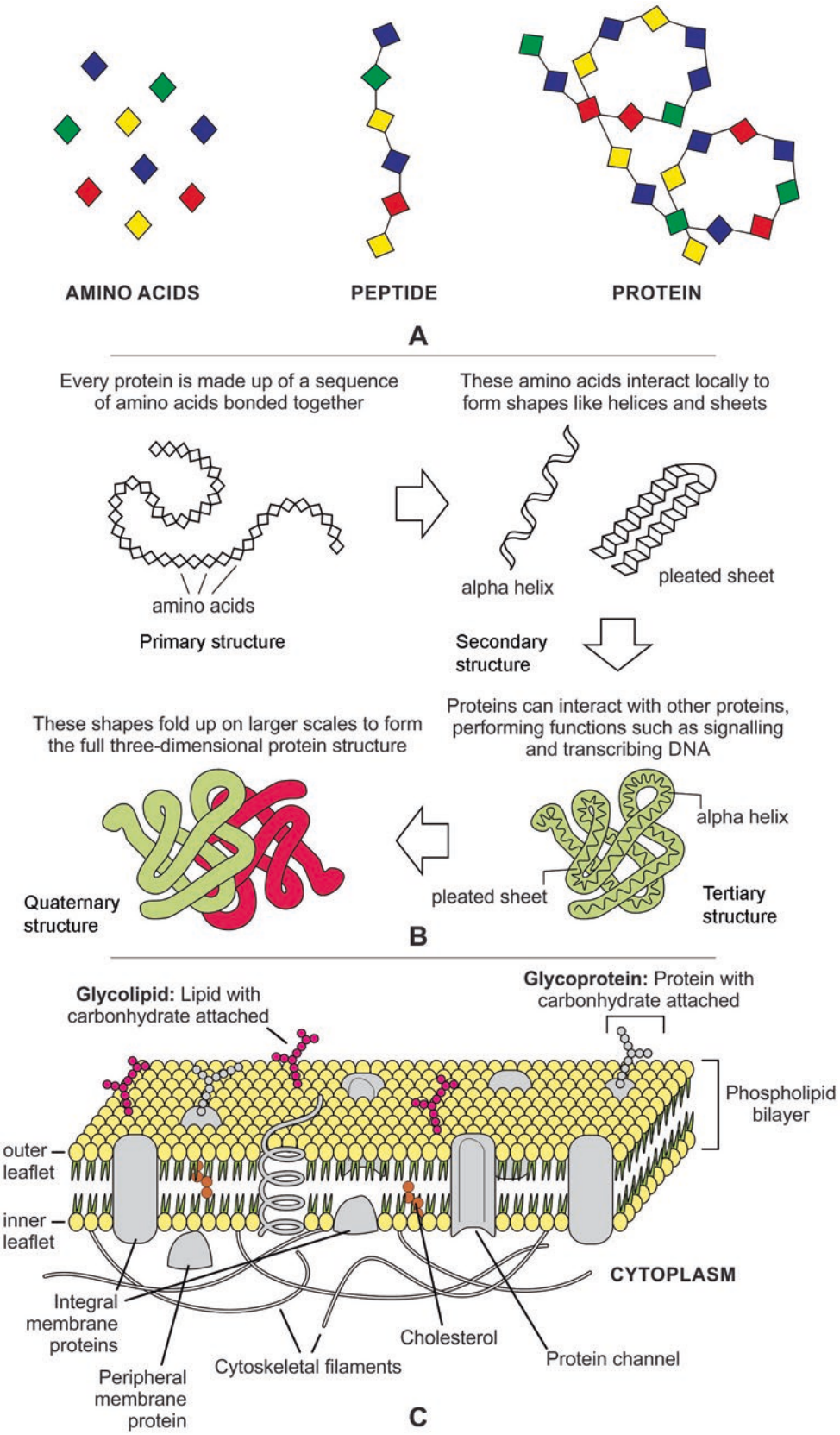


Fig. 14.2 (**a**, **b**) Protein folding. (**c**) A plasma membrane. (**a**) Amino acids joined by peptide bonds to form a peptide and a polypeptide by a condensation reaction. (**b**) Different structures of proteins, from primary to quaternary, are shown by the arrows; proteins have unique amino acid sequences that are specified by genes; the 'primary' structure of a protein—its amino acid sequence—drives the folding and intramolecular binding of the linear amino acid chain, which ultimately determines the protein's unique three-dimensional shape. The hydrogen bond between amino groups and carboxyl groups in the protein chain causes specific folding patterns to occur. The stable folding patterns of alpha-helices and beta-sheets make up the two-dimensional 'secondary' structure of proteins. In enzymes, the long chain of amino acids becomes folded into a three-dimensional functional shape or a 'tertiary' structure. In antibody molecules and hemoglobin, several polypeptides may be bound together to form a 'quaternary' structure. (**c**) A plasma membrane structure showing a phospholipid bilayer. Located throughout the plasma membrane's surface are cholesterol molecules that stabilize the phospholipids and keep them in position. The plasma membrane is also intruded by proteins. Peripheral proteins go halfway through the membrane; others except integral proteins go entirely through the membrane. These proteins serve as channels to allow the molecules to enter and leave the cell

molecules. These special shapes help enzymes lock onto other molecules of complementary configurations so that they fit together like jigsaw puzzle pieces. The fit is more complicated than two puzzle pieces. The 'hand-in-glove' analogy can be used to talk about the flexibility of enzymes; here, an enzyme changes shape to some degree when it binds its substrate, thus resulting in an even tighter fit. Enzymes are flexible and dynamic and can undergo a significant change in shape.

A pocket or a groove is an active site, which is continuously reshaped by interactions with the substrate as it interacts with the enzyme. The substrates fit into the enzyme and then react. What kind of analog information is present between the substrate and enzyme that allows mutual recognition? An analog memory helps find the enzyme–substrate attraction. An enzyme attaches to one or more reactant molecules while catalyzing a reaction in an active site. The bonding is only temporary, but it is enough for the enzymes to bring their targets together in just the right way to ignite a chemical reaction. Enzymes lower a reaction's activation energy—that is, the amount of energy that must be put in for the reaction to begin. They catalyze metabolic reactions by tightly binding to specific substrates and holding them in close juxtaposition to reduce the activation energy. Binding pockets achieve specificity with the complementary shape, charge, and hydrophilic/hydrophobic characteristics of the substrates (Fig. 14.3). Bound to some enzymes is an additional component called a cofactor, which is a direct participant in the catalytic event and is thus required for enzymatic activity.

An enzyme binds its substrate to form an enzyme–substrate complex. The specific molecular geometry in the substrate is recognized by the specific geometry at the active site of the enzyme to obtain the conformational fit. There may be different kinds of substrates for each type of enzyme. Sometimes, a single-reactant substrate breaks down into multiple products. At other times, two substrates may fuse to create a larger molecule. Two reactants may get modified after the reaction, creating two products. The enzyme–substrate recognition system is an example of spatial information, i.e., an analog memory system by molecular shape recognition.

Environmental conditions can affect an enzyme's active site and, therefore, the rate at which a chemical reaction can proceed. Active sites depend on changes in the enzyme's environment, including temperature and pH. Dramatic changes to the temperature and pH will eventually cause enzymes to denature.

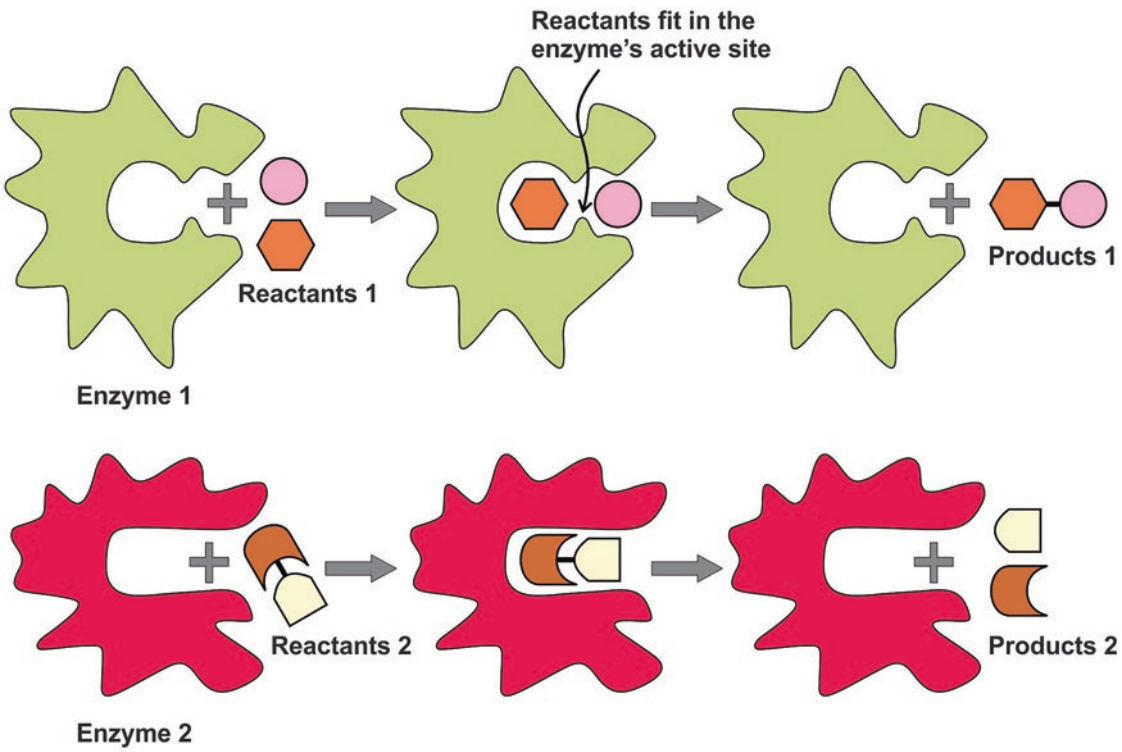


Fig. 14.3 Shape recognition in the enzyme–substrate complex involves analog memory. Different enzymes perform different tasks because they have different shapes. Each enzyme shape accommodates a particular shape of a chemical reactant into an active site, where a chemical reaction occurs to yield products. Top, two substrates may come together to create a larger molecule; bottom, two reactants might enter a reaction, get modified, and leave the reaction as two products

14.3 Protein/RNA World

Proteins are the pathfinders of emerging protocells. They are among the most abundant and diverse molecules in living systems. Undoubtedly, the advent of proteins utterly changed the conditions of emerging life and enhanced abiotic synthesis. They were undoubtedly the product of RNA activity in the peptide/RNA world. In the prebiotic environment, two significant events followed the emergence of mRNA directed proteins in protocells. These are as follows:

1. The evolution of the phospholipid and plasma membranes
2. The protein-rich cytoplasm of protocells

The newly synthesized protein enzymes helped catalyze and mediate these critical molecular evolutions, favored by strong selective forces.

The availability of proteins would open a new world of viruses in parallel with protocells. The primitive virus particle was a hybrid molecule of mRNA with a coat of protein capsid that offered stability and durability in the hydrothermal vent environment.

14.4 The Evolution of Phospholipid Membranes

Most likely, phospholipids did not become available in the prebiotic environment until various protein enzymes were available and metabolic pathways for their catalyzed synthesis had evolved. Phospholipid bilayers were impermeable to most water-soluble molecules. This property made bilayers function as sharp boundaries that allowed for protocells to maintain an internal composition different from that of the surrounding medium. However, protocells could not survive and evolve if sealed off from the outside. They must be able to take up nutrients, get rid of waste products, and respond to environmental signals. These functions were carried out by proteins inserted into phospholipid bilayers to form the plasma membrane, improving the bilayer's permeability.

14.4.1 The Origin of the Phospholipid Membrane

Three critical molecules—fatty acids, glycerol, and a phosphate ion—were available in the hydrothermal vent environment to synthesize phosphatidic acid by nonenzymatic synthesis (Fig. 14.4a, b) [4]. Abiotic synthesis of aliphatic lipids, fatty acids, and acylglycerols has been reported to occur at elevated temperatures and pressures under simulated hydrothermal conditions [5]. A phospholipid is synthesized by a series of enzyme-catalyzed energy-dependent reactions from phosphatidic acid (Fig. 14.4c). This synthesis was possible when RNA-directed protein enzymes were available to convert the phosphatidic acid into a phospholipid. As such, a phospholipid provides a marked contrast to fatty acid as a membrane component. Because of its molecular complexity, it is generally believed that a phospholipid's prebiotic synthesis was difficult [6]. More recently, several studies have achieved the integration of phospholipids and related compounds, such as acylglycerol and glycerol phosphates, suggesting that such molecules may have been present in the prebiotic environment in trace quantities [4, 6–12]. Moreover, the abiotic formation of the ester bond among lipid compounds, including acyl glycerides, is possible under simulated hydrothermal conditions, provided the precursors present are at sufficient concentrations [13].

The synthesis of phospholipids requires an activated intermediate, phosphatidic acid (diglycerol-3-phosphate). The prebiotic pathway from fatty acids to the simplest phospholipid, phosphatidic acid, occurs via successive acyl- and phosphotransfer reactions. The first step in the synthesis of phospholipids is the synthesis of phosphatidic acid, which is formed by adding two fatty acids to glycerol-3-phosphate. Here, we explore the nonenzymatic pathways of the emergence of the phosphatidic acid in the peptide/RNA world. From three building blocks—fatty acids, glycerol, and a phosphate ion, which were available in the prebiotic environment [13], the gradual evolution of monoglycerides, diglycerides, and triglycerides occurred by condensation reactions (Fig. 14.4b).

When glycerol and fatty acids react, a water molecule is expelled, forming an ester linkage. The production of diglyceride is considered first since it will directly lead to the biosynthesis of phosphatidic acid. Diglyceride is formed when the glycerol and fatty acid chains are joined by two ester linkages. Nonenzymatic synthesis of ester bonds to produce diglycerides might have been the first step toward glycerolipids [4, 12, 13]. The phosphorylation of diglyceride, in turn, would give rise to phosphatidic acid, which is essentially a diglyceride in which a phosphate group has been added to a single glycerol molecule (Fig. 14.4c).

The nonenzymatic synthesis of activated phosphatidic acid was a pivotal point in the lipid biosynthetic pathway [14]. It served as the precursor to the formation of the glycerophospholipid (commonly called phospholipid) membrane by enzymatic synthesis. Glycerophospholipids are the main constituents of membrane bilayers. Enzymatic synthesis pathways evolved over time when RNA-directed protein enzymes were available in the protein/RNA world. The phosphate group of activated phosphatidic acid is esterified to alcohol to produce a variety of phospholipids, including the attachment of choline, ethanolamine, serine, and inositol

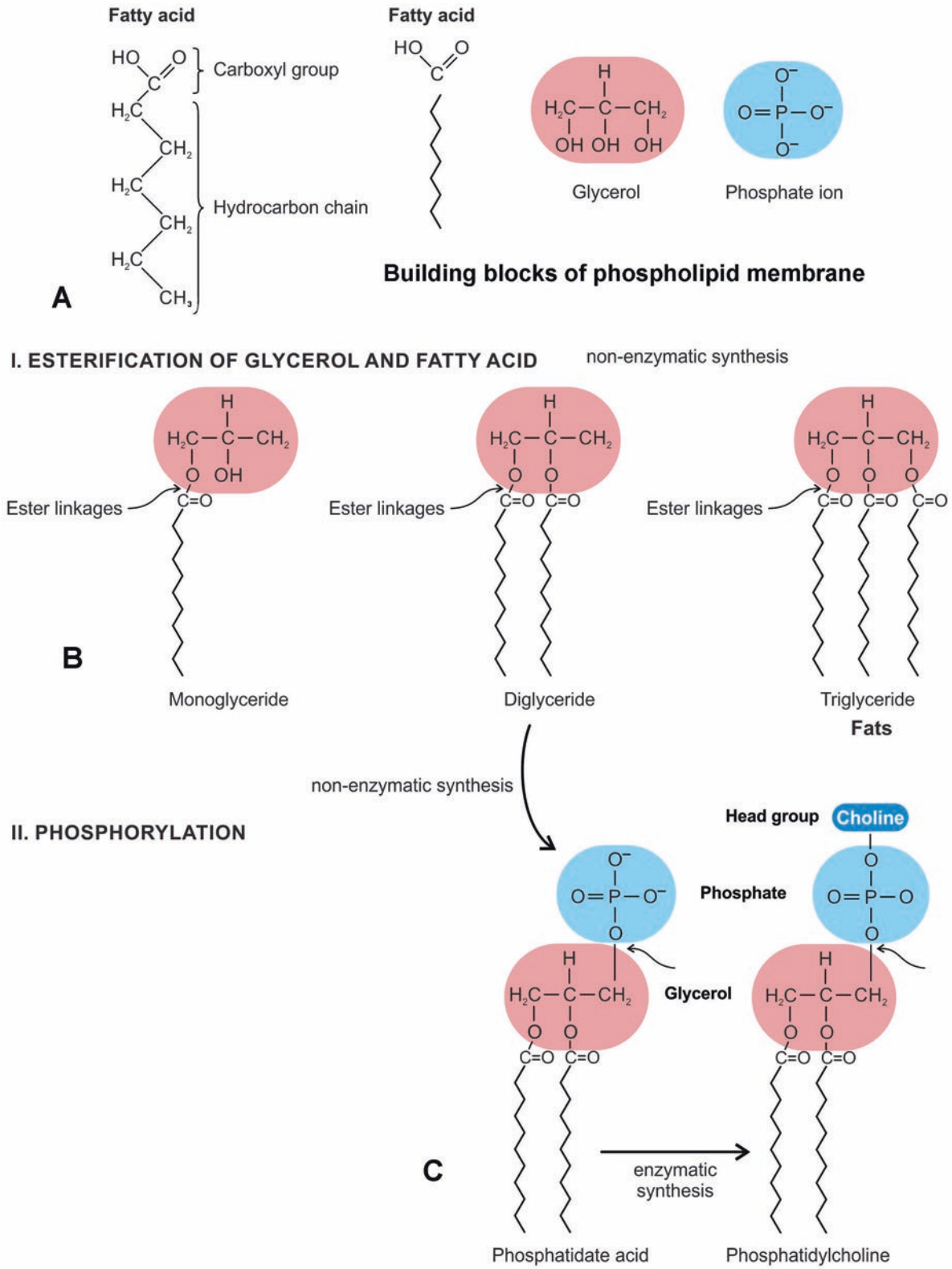


Fig. 14.4 The origin of the phospholipid membrane from simple fatty acids by an intermediate, the phosphatidic acid, by a series of nonenzymatic syntheses. (**a**) Fatty acids, glycerol, and a phosphate ion are the building blocks of phospholipids. (**b**) In the first stage of phospholipid formation, several glycerides such as monoglycerides, diglycerides, and triglycerides (fats) are formed by esterification of glycerol and fatty acids, with the loss of a water molecule; the covalent bond, an ester linkage, results from this reaction. (**c**) The next stage of synthesis of phosphatidic acid is by phosphorylation of a diglyceride molecule when a phosphate ion is joined. In turn, phosphatidic acid would give rise to a phospholipid by attaching to an alcohol molecule such as choline, ethanolamine, serine, or inositol. Of these various combinations, phosphatidylcholine (shown in the figure) is the most common phospholipid in the cell membrane

to the phosphate group of phosphatidic acid. These phospholipids then include phosphatidylcholine (phosphate + choline), phosphatidylethanolamine (phosphate + ethanolamine), phosphatidylserine (phosphate + serine), and phosphatidylinositol (phosphate + inositol). If the alcohol is choline, then the product is phosphatidylcholine. Of these, phosphatidylethanolamine is the most common phospholipid in bacterial cell membranes. Different modifiers give the phospholipids different properties and roles in a cell. Three successive enzymatic methylations could convert the phosphatidic acid into a phospholipid.

Phosphate is the primary anionic component of most phospholipid membrane lipids. A phospholipid consists of a polar head group on one end of the molecule and fatty acid chains on the other end. These chemical structures create an amphipathic liquid. In solution, they instantly form bilayers that are selectively permeable. Phospholipids are composed of a polar head group (usually a negatively charged phosphate group and glycerol); it is hydrophilic. A phospholipid's tails consist of two long fatty acid chains, which are hydrophobic and avoid interactions with water. Two fatty acids are attached to glycerol by esters or other bonds (Fig. 14.4a). The polar head group and fatty acid chains are attached by a three-carbon glycerol unit [6].

Because phospholipid molecules have both a hydrophilic and hydrophobic group on the same molecule, they can undergo self-assembly into a cell. The hydrophobic tails mix with oil in an oil slick, whereas the heads stay close to the water in a monolayer cell (or micelle). When placed in water, the phospholipids will orient themselves in a bilayer in which nonpolar tail regions face the inner layer of the bilayer (Fig. 14.5b). Phospholipid molecules, because of their cylindrical structure, provide structural stability and create a semipermeable environment. The same forces that drive the phospholipids to form bilayers also provide a self-healing property. The admixture of cholesterol helps stabilize the bilayer.

Modern phospholipid-based cell membranes are formidable barriers to the uptake of polar and charged molecules ranging from metal ions to complex nutrients and nucleotides. They require special protein transporters to allow their passage through the membrane. Phospholipid membranes are stable under a wide range of temperatures, pH, and salt concentrations (Fig. 12.3). A recent experiment has suggested

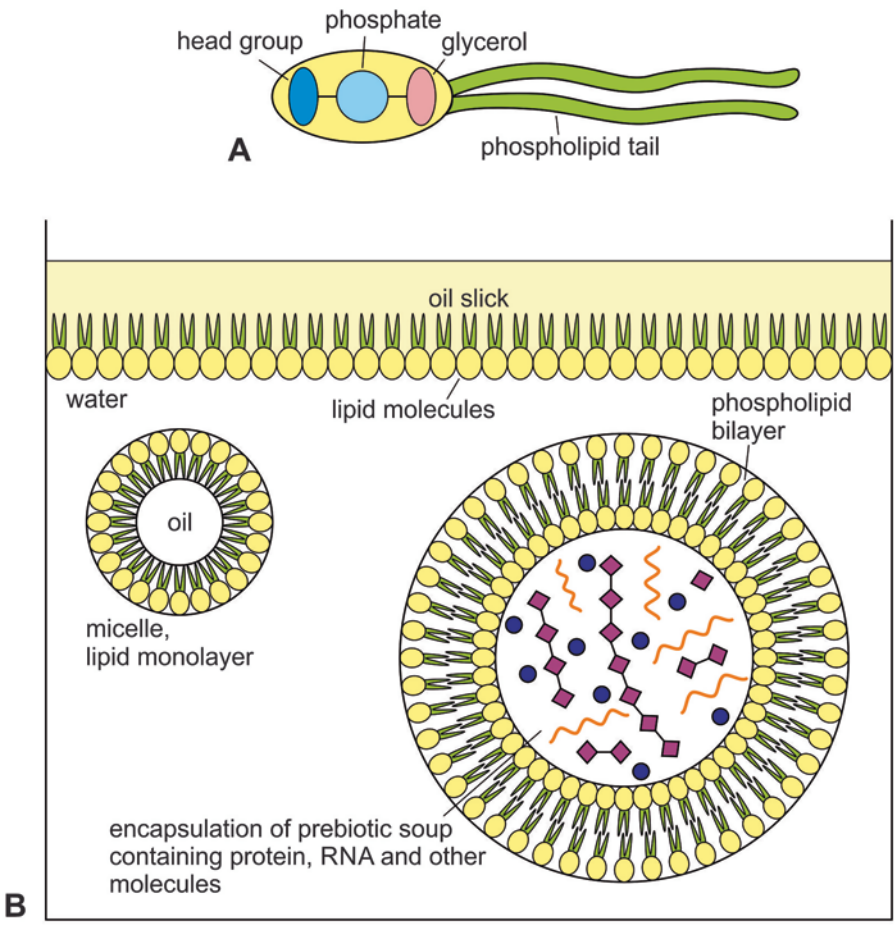


Fig. 14.5 Self-assembly of a phospholipid membrane in a hydrothermal crater lake in the protein/RNA world. (**a**) Generalized phospholipid molecule has a hydrophilic ('water-loving') head and two hydrophobic ('water-hating') tails that do not mix with water and avoid being surrounded by it. (**b**) In an oil slick on the surface of a crater lake, the hydrophobic tails mix with oil, whereas the heads stay close to water. During turbulence, phospholipids form two kinds of membranes: a monolayer, which can only capture a drop of oil (left), or a bilayer, which can capture a group of water molecules (right). The bilayer allows the hydrophobic tails to associate with one another; the heads are associated with water molecules on both the inside and outside of the membrane. A bilayer vesicle is stabilized when it encapsulates protein molecules that interact with the bilayer surface

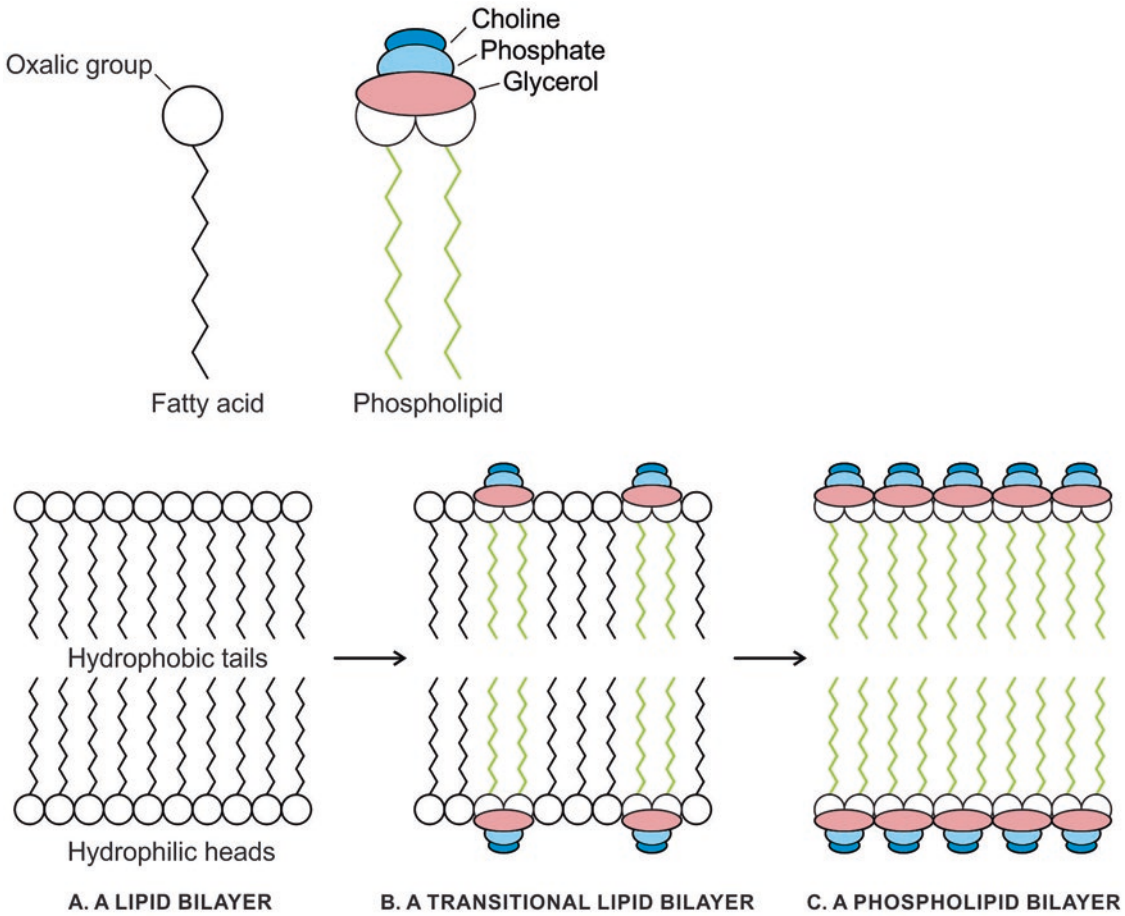


Fig. 14.6 The gradual transition from a single-chain, highly permeable fatty acid membrane to a selectively permeable phospholipid membrane, with an increase of phospholipid content via a transitional stage. Increasing phospholipid content inhibits the permeability of fatty acid membranes through changes in bilayer fluidity

that phospholipid membranes can self-assemble and are sufficiently permeable to transport molecules across their membranes [15].

The evolution of phospholipid membranes was a critical and necessary step for the early evolution of cell membranes. The transition from single-chain lipids to double-chain phospholipids had to be gradual. Recent experiments have suggested that low phospholipids could drive the growth of fatty acid vesicles by competing with adjacent non-phospholipid vesicles. This competitive growth would have provided a strong selective advantage for primitive cells to evolve the catalytic machinery needed to synthesize phospholipids from their single-chain precursors. Growth results from decreasing fatty efflux from the membrane with increasing phospholipid content, suggesting an evolutionary arms race among primitive protocells [8]. The gradual transition from a fatty acid bilayer membrane to a phospholipid bilayer membrane is shown here in Fig. 14.6.

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

14.4.2 The Plasma Membrane

The lipid bilayer is an inert, impermeable membrane that can neither mediate nor control the exchanges of matter across the membrane as well as other functions. These functions are accomplished by integral membrane proteins embedded in the bilayer. Another innovation with the availability of proteins in the prebiotic environment was the emergence of the plasma membrane from the phospholipid membrane, found in all living cells, which separates the interior of the cell from the outside environment in a controlled manner. Cells must be able to exclude, take in, and excrete various substances, all in specific amounts. Besides, they must be able to communicate with other cells, identify themselves, and share

information. Functionally, these proteins include transporters, which facilitate the passive diffusion of their substrates, and active transport systems, or pumps, which can move their substrates forcibly against a concentration gradient, with the help of ATP or some other sources of energy.

Structure of the Plasma Membrane

Plasma membranes are the selectively permeable outermost structures of cells, which separate the cells' interior from the environment. The main components of the plasma membrane are lipids (phospholipids and cholesterol), proteins, and carbohydrates. The plasma membrane is about 50% lipid by weight and almost 50% protein. The plasma membrane's basic structure is a phospholipid bilayer associated with proteins that are largely responsible for many functions of the membrane. There are three types of membrane proteins, including integral, peripheral, and lipid-anchored proteins. Two antiparallel sheets of phospholipids form the membrane that surrounds the content of the cell. The layer closest to the cytoplasm is the inner leaflet, whereas the layer closest to the exterior environment is the outer leaflet. While some proteins span the membrane with structures that cross from one side to the other, others are anchored to membrane lipids, and, still, others are associated with the cytosolic side of a plasma membrane. There are three classes of amphiphilic lipids in the cell membrane: phospholipids, cholesterol, and glycolipids, of which phospholipids contribute more than 50% of all lipids. The third major ingredient of plasma membranes is carbohydrates. In general, they are present only on the plasma membrane's outer surface and are attached to proteins, forming glycoproteins, or to lipids, forming glycolipids (Fig. 14.2c).

Membrane Transport

The plasma membrane of a cell is selectively permeable, in that the membrane's protein channels will only transport certain kinds of molecules across the membrane. There is a formidable array of gates, checkpoints, transport systems, sensors, and antennas that control molecules' transport. Proteins embedded in the plasma membrane are important for facilitating the transport of ions and other molecules in cells. These proteins are ion channels, transporters, and pumps. Most ions and molecules pass through cell membranes via either passive transport—in which the movement is down the gradient and thus requires no expenditure by the cell—or active transport, in which the movement is against the gradient, which requires energy. Most membranes use passive transport, where molecules are moved across the plasma membrane along, down, or with their concentration gradients. If a molecule can cross a plasma membrane, then it will diffuse across its membranes until it reaches equilibrium. The molecule flows from higher to lower concentrations.

In contrast, in active transport, molecules are moved across their concentration gradient in energy-requiring processes. During active transport, carrier proteins pick up the molecule to be transported on one side of the membrane and, with the help of ATP, change shape to move the molecule to the other side of the membrane. Carrier proteins that carry out active transport are called pumps. Active transport is among the most important of cell functions. It is necessary to move certain sugars and amino acids into cells. Nucleotides used in the formation of DNA enter cells via active transport.

Origin of the Plasma Membrane

Here, we suggest the likely scenario for the origin of the plasma membrane from the phospholipid membrane (Fig. 14.7). A new class of proteins emerged in the protein/RNA world, which played critical roles in converting the phospholipid membrane to the plasma membrane. Proteins could be amphiphilic because they are made of amino acids, and amino acids have R-groups that range from highly nonpolar to highly polar. Nonpolar amino acids would be selected for incorporation in the phospholipid membrane in the interior of the lipid bilayer, whereas polar amino acids would be selected alongside the polar heads of the surrounding water. The peptide channel across the lipid bilayer membrane could have been a precursor to insertion of proteins into phospholipid membrane to enhance permeability of the plasma membrane (see Fig. 9.3).

As the phospholipid membrane began to interact with newly generated amphiphilic proteins, endosymbiosis in protocells would integrate these proteins the phospholipid membrane (Fig. 14.7a). These protein symbionts would stabilize the walls of the membrane, allowing them to resist disruptive forces such as the mechanical shear caused by the vent's convection current. The resulting increase in osmotic pressure and membrane tension would create a driving force for growth in the membrane area. Therefore, it would encourage symbiosis between proteins and lipid membranes that triggered the origin of the primitive plasma membrane for stability to prevent a burst of protocells (Fig. 14.7b). More sophisticated plasma membranes might have been formed by endosymbiosis to make them more permeable. The plasma membrane of modern cells is composed of roughly equal parts of proteins and lipids by weight. Most likely, the primitive protocells had a higher percentage of lipids than proteins from the initial endosymbiosis that favors a high degree of thermostability [16]. The plasma membrane acted as a selectively permeable barrier, preventing some substances from crossing while permitting other substances to enter and leave the protocell. The selective permeability of the plasma membrane and the specificity of transport proteins made it possible to create an environment inside the protocell that was radically different from the prebiotic soup and amenable to biogenesis.

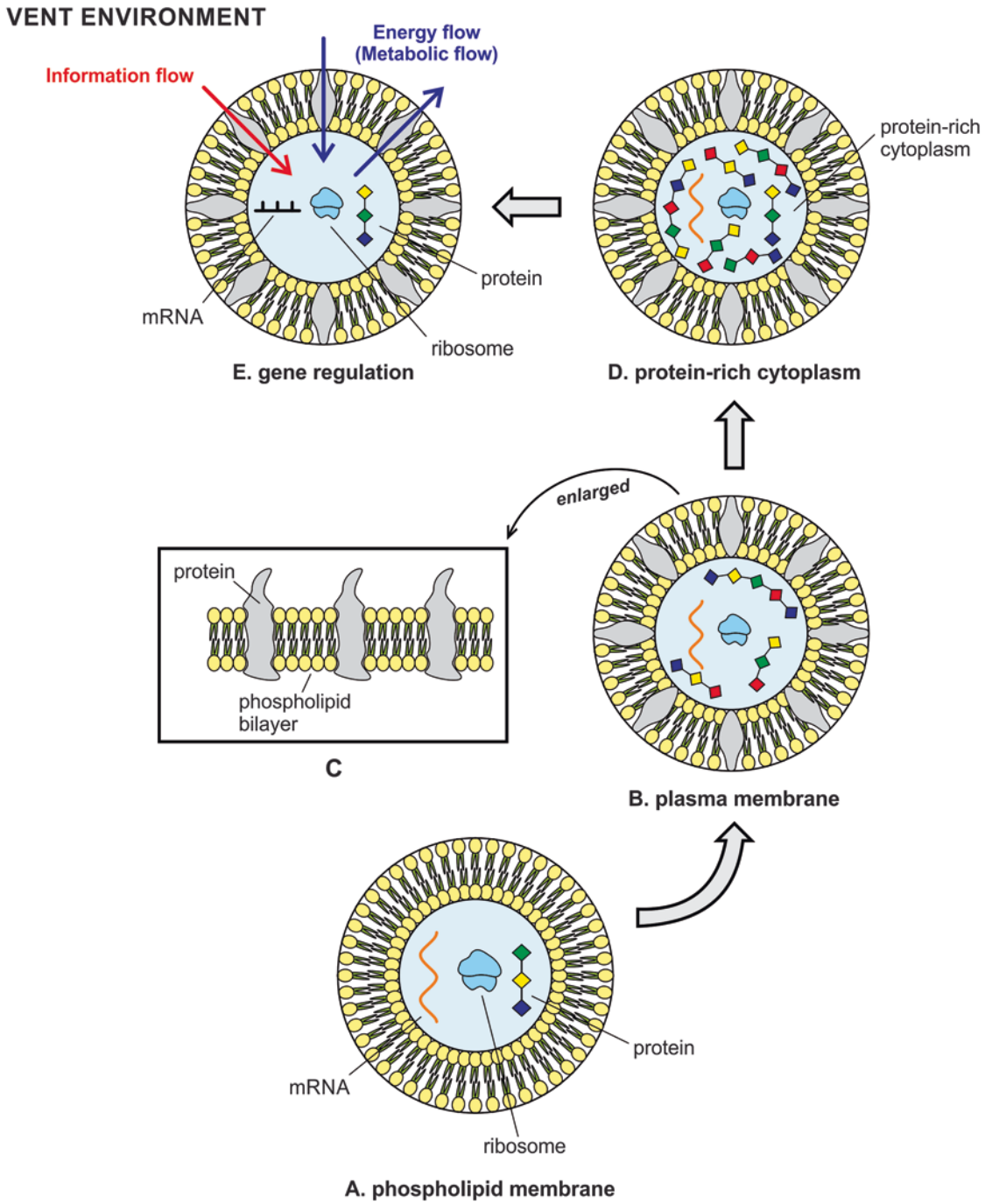


Fig. 14.7 The transition of the phospholipid membrane to the plasma membrane. (**a**) A phospholipid membrane. (**b**) The plasma membrane. (**c**) The phospholipid membrane evolved into the plasma membrane by inserting protein molecules into bilayers that made the cell membrane selectively permeable so that specific ions like potassium and sodium can cross the bilayer barrier. (**d**) The primitive cytoplasm inside the protocell became enriched when newly synthesized protein molecules by the translation machine began to concentrate inside the protocell, thus turning into a gel-like substance. (**e**) Gene regulation in the protein/RNA protocell, energy flow, and information flow are represented as the essential exchanges of the protocell with the outer/inner environment. The protein/RNA protocell began to respond to environmental cues and target mRNA for regulating gene expression

The first step of survival in confinement was the possibility for the protocells to take in nutrients and energy from the outside environment and get rid of waste material. The simplest way fully enveloped protocells could fulfill this condition was utilizing pores, little holes kept open in the phospholipid bilayers by inserting a protein framework. The insertion of proteins into bilayer lipid vesicles was an essential first step that affected the phospholipid bilayer's

permeability, facilitating transport and other molecules into the protocells. In primitive membranes, passive transport was used when molecules were moved across the plasma membrane of the protocells along, down, or with their concentration gradients. The molecules would flow from a higher concentration to a lower concentration.

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

The next improvement in the building of molecular transport systems was active transport, in which molecules were moved against their concentration gradient in energy-requiring processes. The machinery involved was correspondingly more complex. The systems that carry out such active transport are called pumps. The energy used was derived from ATP, the universal currency of energy. Various ATP-powered pumps were used to transport ions and molecules against their concentration gradients. The plasma membrane would have played an essential role in the generation of metabolic energy and transforming it into useful ATP.

The fluid mosaic model of the cell membrane structure distinguishes between two types of membrane proteins: peripheral proteins and integral proteins; the former occurs only outside the lipid bilayer, whereas the latter spans the entire membrane to transport ions and molecules [17]. The phospholipids and the plasma membranes make back and forth movements within the plasma membrane, making the plasma membrane a fluid structure. The critical point is that the arrangement of proteins makes the interior and exterior surfaces of the plasma membrane extremely different (Fig. 14.2c). Thus, the plasma membranes are a mosaic of phospholipids and the different proteins, and the overall structure is dynamic and fluid. In protocells, the membrane proteins were responsible for the passage of ions, polar molecules, and large molecules that did not readily cross the phospholipid bilayers independently.

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

Proteins are responsible for most of the dynamic processes carried out by the cell membrane, including the transport of molecules in and out of the cell (Figs. 14.2c and 14.7e). The plasma membrane separates the cell from its environment and is selectively permeable: it chooses what enters and exists the cell. Receptor proteins are the gatekeepers; they detect signals from the environment of the cells; transport proteins help some molecules get across the membrane. Certain membrane proteins act as enzymes. The plasma membrane was the ideal microenvironment to experiment with the synthesis of more complex nucleic acids, such as DNA, for the permanent storage of the information system.

14.5 The Protein-Rich Cytoplasm of Protocells

We had previously discussed how the primitive protocell enclosed assemblages of peptides and nonenzymatic RNA molecules in the primitive cytoplasm (see Fig. 9.3). Once different kinds of proteins were synthesized and available for prebiotic synthesis, they began to accumulate in the cytoplasm inside protocells. This primordial cytoplasm became the ready source of a variety of proteins when needed. The primitive cytoplasm inside the protocell was gradually converted into a viscous, gel-like cytoplasm that increased the protocell volume and provided some rigidity to its spherical shape. The protein-rich cytoplasm in protocells and the semipermeable plasma membrane from the phospholipid membrane most likely organized at the same time, with the availability of proteins (Fig. 14.7d). The prebiotic cytoplasm provided a stable microenvironment for the organization of all nucleic acids, lipids, enzymes, proteins, other macromolecules, molecules, ions, and ribosomes as well as water and salts that were all encased in the plasma membrane. The cytoplasm offers a safe heaven for the essential molecules and machinery that can be sequestered and self-contained from the external environment to maintain homeostasis, except for the signals, or receptors. The protocell used recursive procedure when certain molecule or machinery needed from the storehouse of cytoplasm to perform any specific function. This primitive cytoplasm became a complex, crowded system containing a wide range of molecules—from ions and small molecules to macromolecules like proteins, nucleic acids, and ribosomes.

Many metabolic reactions, including protein synthesis and the transition from RNA to DNA, occurred in this primordial cytoplasm. Over half of the molecules were actively involved in the synthesis of proteins. Some of these proteins were used in the synthesis of viruses and DNA molecules, whereas others were engaged in energy production. The constituents of the cytoplasm moved across the protocell, depending on their requirements. Primitive cytoplasm supported and suspended these molecules in their gel-like substance. This primordial cytoplasm became the site for most of the enzymatic reactions and metabolic activity of the protocell. The primitive cytoplasm was confined to the outside by the plasma membrane; the latter began to regulate the passage of some substances, such as organic molecules, ions, and water, preventing the passage of some other substances to maintain the content of the primitive cytoplasm. Other compounds moved passively across the membrane.

14.6 Gene Regulation of Protocells and Information Flow

mRNA-directed protein synthesis was the first stage of gene expression. mRNA protocells were open systems regarding both energy flow and information flow. Another creative outcome of protein synthesis was the development of a rudimentary gene regulatory system. It consisted of protocells choosing which mRNA gene to use and which not to use during protein synthesis from the environmental signal. In general, a gene was expressed only when its protein product was needed. RNA protocells responded according to environmental fluctuations, transmitted signals across its plasma membrane, and targeted specific mRNA for protein synthesis. Plasma membranes developed various receptor proteins to control information flow from the outer environment to mRNAs in the cytoplasm. These mRNA genes were mainly constitutive genes that contained the blueprints for proteins for essential housekeeping functions.

Gene expression occurred at translational and posttranslational levels (Fig. 14.7e). Perhaps a regulatory protein bound a specific mRNA in response to environmental fluctuation and increased the translational rate to build a corresponding protein. These proteins were essential for the survival of the protocells. In general, each mRNA molecule went on to make a specific protein. In some cases, this protein would be an enzyme; in others, it would be structural, according to the protocell's demand.

Some proteins were needed for basic metabolisms, such as the enzymes that catalyzed reactions during glycolysis so that protocells could transfer energy from food to ATP. Others would be required to make the viruses, the protective coat of mRNA, which would, in turn, give rise to DNA. Proteins were not fully formed and functional when termination from ribosomes occurred. They were manufactured in an inactive form and then had to be activated by chemical modification, such as the addition of a phosphorous group. This type of change is posttranslational control.

The following steps occurred as information flowed from the environment to mRNA to proteins represented by arrows in the following expression:

Environmental signal → mRNA → protein → activated protein.

14.7 Conclusions

Translation of mRNA to proteins underpins the molecular symbiosis that has dominated life for four billion years. This universal unity of digital information provides a blueprint of the common origin of all living systems. The emergence of proteins in the prebiotic world was a turning point in the origin of life. With their astonishing versatility, the protein enzymes catalyzed crucial biochemical reactions within protocells into more complex biomolecules in diverse metabolic pathways. In contrast, structural proteins provided strength and permeability to cell membranes. During translation, each protein was synthesized as a linear chain of amino acids on ribosomes. The ordering of the amino acid chain determined how the protein chain would fold upon itself to become a biologically active protein in its native 3D structure.

Three major biochemical innovations followed in succession after the availability of various kinds of protein molecules during prebiotic synthesis. These are (1) the modification of the phospholipid membrane into the plasma membrane; (2) the enrichment of the primitive cytoplasm with protein enzymes; and (3) primitive gene regulation. Protein enzymes significantly speeded up the prebiotic chemical reaction. Some enzymes helped bind two molecules together to produce a new molecule. Others helped break large molecules into smaller pieces. The inherent information system in proteins is not properly investigated. It may provide a new window to prebiotic synthesis.

Proteins also helped modify the bilayer membrane into a phospholipid membrane and then the plasma membrane. Plasma membrane stabilized the wall of the protocell and made it more permeable, preventing some molecules from the protocells from crossing the membrane barriers while permitting other selected molecules and ions to enter and leave the protocell. With the availability of proteins, there was a gradual conversion of the interior of the protocell from a water-like medium into a gel-like cytoplasm, which became the storehouse of a wide range of biomolecules, including

amino acids, proteins, nucleic acids, and ribosomes, as well as salt and water. Plasma membrane facilitated gene regulation and information flow from the environment to mRNA for specific protein synthesis. The access of new generation of proteins would encourage the rise of viruses in the hydrothermal vent environment.

References

1. De Duve C (2005) Singularities: landmarks on the pathways of life. Cambridge University Press, New York
2. Osawa S (1995) Evolution of the genetic code. Oxford University Press, Oxford
3. Freeman S (2005) Biological science, 2nd edn. Pearson Prentice Hall, Upper Saddle River
4. Hargraves WW, Mulvihill SJ, Deamer DW (1977) Synthesis of phospholipids and membranes in prebiotic conditions. Nature 266:78–80
5. Deamer DW (2017) The role of lipid membrane in life's origin. Life 7:5. https://doi.org/10.3390/life3701005
6. Sandford SA, Bernstein MP, Dworkin JP (2001) Assessment of the interstellar processes leading to deuterium enrichment in meteoritic organics. Meteorit Planet Sci 36:1117–1133
7. Monnard PA, Appel CL, Kanavarioti A, Deamer DW (2002) Influence of ionic inorganic solutes on self-assembly and polymerization processes related to early form of life: implications for a prebiotic aqueous medium. Astrobiology 2:139–152
8. Budin I, Prwyes N, Zhang N, Szostak JW (2014) Chain-length heterogeneity allows for assembly of fatty acid vesicles in dilute solutions. Biophys J 107. https://doi.org/10.1016/j.pbj.2014.07.067
9. Zhu TF, Szostak JW (2009) Coupled growth and division of model protocell membranes. J Am Chem Soc 131:5705–5713
10. Chen IA, Szostak JW (2004) A kinetic study of the growth of fatty acid vesicles. Biophys J 87:988–998
11. Budin I, Szostak JW (2011) Physical effects underlying the transition from primitive to modern cell membranes. Proc Natl Acad Sci 108:549–5254. https://doi.org/10.1093/molbev/msx265
12. Apel CL, Deamer D (2005) The formation of glycerol monodecanoate by a dehydration condensation reaction: the chemical complexity of amphiphiles on the early Earth. Orig Life Evol Biosph 35:323–332
13. Simoneit BRT, Rushdi AI, Deamer DW (2007) Abiotic formation of acylglycerols under simulated hydrothermal conditions and self-assembly properties of such lipid products. Adv Space Res 40:1649–1656
14. Deamer DW (2011) Self-assembly of organic molecules and the origin of cellular life
15. Mansy SS, Schrum IP, Krishnamurthy M, Tobe S, Treco DA, Szostak JW (2008) Template-directed synthesis of a genetic polymer in a model protocell. Nature 454:122–126
16. Ross DS, Deamer D (2016) Dry/wet cycling and the thermodynamics and kinetics of prebiotic polymer synthesis. Life 6:28. https://doi.org/10.3390/life6030028
17. Singer SJ, Nicholson GL (1972) The fluid mosaic model of the structure of the cell membrane. Science 175:720–731

The Virus World in Deep Time

15

Viruses are completely selfish. They break into cells, overpower their normal functioning, and coerce them into one task: the production of more viruses.

—David S. Goodsell, 2009

Viruses are tiny, noncellular, microscopic parasites that infect virtually every type of known organism. Even simpler and smaller than a bacterium, a virus has a diameter of 20–400 nanometers (nm), one billionth of a meter. Viruses are everywhere. Wherever there is life, there are viruses! Viruses outnumber stars in the universe. An estimated 10^{32} individual viruses exist on our planet. They infiltrate every nook and corner of our natural world, teeming in seawater, drifting through the atmosphere, lurking in motes of soil, and infecting plants and animals. In extreme environments, they are found from hydrothermal pools to hypersaline lakes, deserts, and several kilometers below and above Earth's surface. Viruses circle the planet in the troposphere, above the planet's weather systems. Trillions upon trillions of viruses fall from the sky each day. The incredible diversity of viruses and their limited genetic complexity suggest that they might have emerged before the first cells' origins. They are the 'relicts' of the primordial protein/RNA world, coevolving with protocells in hydrothermal crater vents.

15.1 Viruses: Pirates of the Cellular World

A virus has a core of genetic material, either RNA or DNA, surrounded by a protective protein coat called a capsid. A virus cannot replicate alone. Viruses must infect cells and use the translation machine of the host cell to make copies of themselves. It is debatable whether viruses are alive since they hijack the cells of living bodies to create more viruses by turning the cells into virus factories. Viruses straddle the line between living and nonliving and between digital and hybrid information systems. Because they are not living organisms in the true sense, they require a cell's biochemical machinery to reproduce. A virus is just a few strands of genetic material wrapped in a package of a protein and lipid overcoat—a parasite unable to function independently. An inert-state virus is a hybrid molecule of nucleic acid with a protein coat. To survive, it must find a cell to infect. Only then can the virus become alive to take control of the host's cellular machinery and use it to churn out thousands of copies of itself. Similar to hybrid car that switches from gas to electric, a virus exists in hybrid form when it is dormant, but switches to digital mode during infection of a host cell to produce more viruses. These viruses then moves from one cell to the next, transforming each new host into a factory that makes even more virus particle.

Living things could use their metabolism to survive, grow, and reproduce. Viruses lack biological machinery to replicate. They do not undergo metabolism. They do not multiply through cellular division. Living things have cells. Viruses do not have cells. They do not even come close to meeting all the standards for life (see Chap. 2). They are at the edge of life. In 2000, the International Committee on Taxonomy of Viruses declared that 'viruses are not living organisms.' Although viruses are not living, they are vital members of the web of life. The ancient evolutionary history of viruses occurs in parallel with that of the cellular world.

Viruses have several hallmarks, including tiny genomes, a small size, and parasitic dependence on cellular hosts for replication. These factors make them unique and set them apart from all other living things despite their animation. However, the discovery of the giant viruses (~600 nm), called the 'mimivirus,' with massive genomes and complete resources for building proteins, further blurs the established boundaries between viruses and the smallest parasitic cellular organisms. The simple size-based distinction between viruses and cells is no longer tenable. However, its icosahedral ultrastructure comprising a capsid coat and its typical eclipse phase in its life cycle support the viral nature of the mimivirus. Furthermore, the mimivirus lacks universal translation machinery, such as encoding ribosomal RNA and proteins [1].

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_15

Viruses can be defined as capsid-encoding organisms as opposed to ribosome-encoding cellular microorganisms [2, 3]. Viral particles are by far the most abundant biological entities on our planet, greatly outnumbering all their cellular hosts; most of the biomass in the ocean is made up of viruses [4]. The genetic diversity of viruses is enormous as well because they can acquire genomes from their hosts, and they can later pass on these genes to their new hosts by horizontal gene transfer. Viruses are agents for gene dissemination, evolution, and biodiversity.

Viruses infect a wide range of organisms, such as bacteria, algae, plants, fungi, insects, and animals, including vertebrates. They are completely selfish and obligate parasites [5]. They cannot independently reproduce because they lack ribosomes and the rest of the living cells' protein-making machinery. They can only reproduce by invading host cells and hijacking their ribosomes, enzymes, and energy. When a virus enters a cell, it sheds its capsid coat, bares its genes, replicates, and induces the cell's translation machinery to manufacture more viral proteins from the viral nucleic acid. The viral genes and capsid protein self-assemble to form virions, turn the host cell into a virus factory, and strain the cell to its bursting point. The host cell explodes and releases hundreds of virus progeny by the lytic cycle. Some viruses are also capable of lysogeny, and their genomes integrate into the host cell chromosome. Thus, viruses only need to do two things: they need to have a mechanism for replication within the host cells and they need a way to get out of their target cells [8].

Viruses are fast to reproduce. A flu virus in humans can generate billions of itself within hours. Moreover, they evolve 10,000 times faster than humans because they have a high mutation rate, and, sometimes, these mutations result in a new variant of the virus. Some variants emerge and disappear, whereas others persist. In our lifetime, we are all aware of the devastating impact of the coronavirus disease 2019 (COVID-19) pandemic that led to a dramatic loss of human life worldwide and presented an unprecedented challenge to public health, societies, and economies. The Centers for Disease Control and Prevention (CDC) and other public organizations currently monitor all variants of the virus that causes COVID-19. One of the variants of concern is Omicron (B.1.1.529, BA.1, BA.1.1, BA2, BA.3, BA.4, and BA.5 lineages) as of July 2022.

Viruses have preferred this parasitic lifestyle for four billion years and are still going strong in their own way. Viruses and cellular life are coevolving in two parallel worlds and maintaining their unique reproductive strategies. Virus-derived sequences constitute a substantial portion of many cellular genomes. All viruses extant today have survived by pirating the host cell machinery. Therefore, each virus has a way of reproducing and expressing its nucleic acid; these strategies exploit the tools available within the susceptible host cell [5].

15.2 The Structure of Viruses: Hybrid Components

A virus is a kind of nanoparticle that is visible only using an electron microscope, and viral exhibits come in an astounding variety of shapes. The simplest viruses have just two hybrid components: a nucleic acid core (digital) and an outer protein capsid shell (analog). The capsid consists of protein subunits called capsomeres and plays a protective role, sequestering the genome from physical and chemical damaging agents. Two kinds of viral capsid structures are found: a cylindrical-shaped helical capsid (Fig. 15.1a) and a spherical icosahedral capsid (Fig. 15.1b). Viruses may also contain additional proteins, such as receptor or attachment proteins, which allow the virus to adhere to the outside of the host cell. The genome, either DNA or RNA, contains instructions for taking over cells, making capsids, and creating more virions or viral particles. The few genes that viruses carried allowed them to execute the minimal tasks required for making new viruses: to invade a cell and slip their genes into the cell's translation machinery. There are many types of viruses, classified on the basis of their size and shape, their genetic material (RNA or DNA), and their host organisms. Most viruses have a genome based on DNA, although a significant minority has RNA genes. Viruses are highly diverse in their morphology and, likewise, in their genetic material. Encapsidation of viral genomes constitutes a virus particle.

In their overall structure, viruses fall into two categories: enveloped and non-enveloped, depending on the presence or absence of an envelope. Non-enveloped viruses have an extremely simple structure. They consist of genetic material and possibly one or more enzymes that are encased in the capsid. Non-enveloped viruses include adenoviruses, rotaviruses, and viruses for polio, foot and mouth disease, cold and cough, and hepatitis A. A non-enveloped virus is aptly called the 'naked virus' because it lacks an envelope [6, 7].

Some human and animal viruses have an additional lipid bilayer membrane wrapped around the capsid of the virus particles. These viruses are called enveloped viruses. The phospholipid bilayer membrane is stolen from the host cell as new viruses bud and exit from the host cell following infection and replication. Enveloped viruses are more evolved and complex than are the non-enveloped type, where the capsid is surrounded by an envelope (Fig. 15.1d). The envelope consists of a phospholipid bilayer with a mixture of viral proteins and proteins derived from the host cell's plasma membrane [6, 7]. The viral envelope offers a virus some advantage over non-enveloped viruses, including better protection from the host's enzymes, certain chemicals, and the immune system. Enveloped proteins can include glycoproteins, which act as receptor molecules. For these viruses, attachment is a requirement for later penetration of the cell

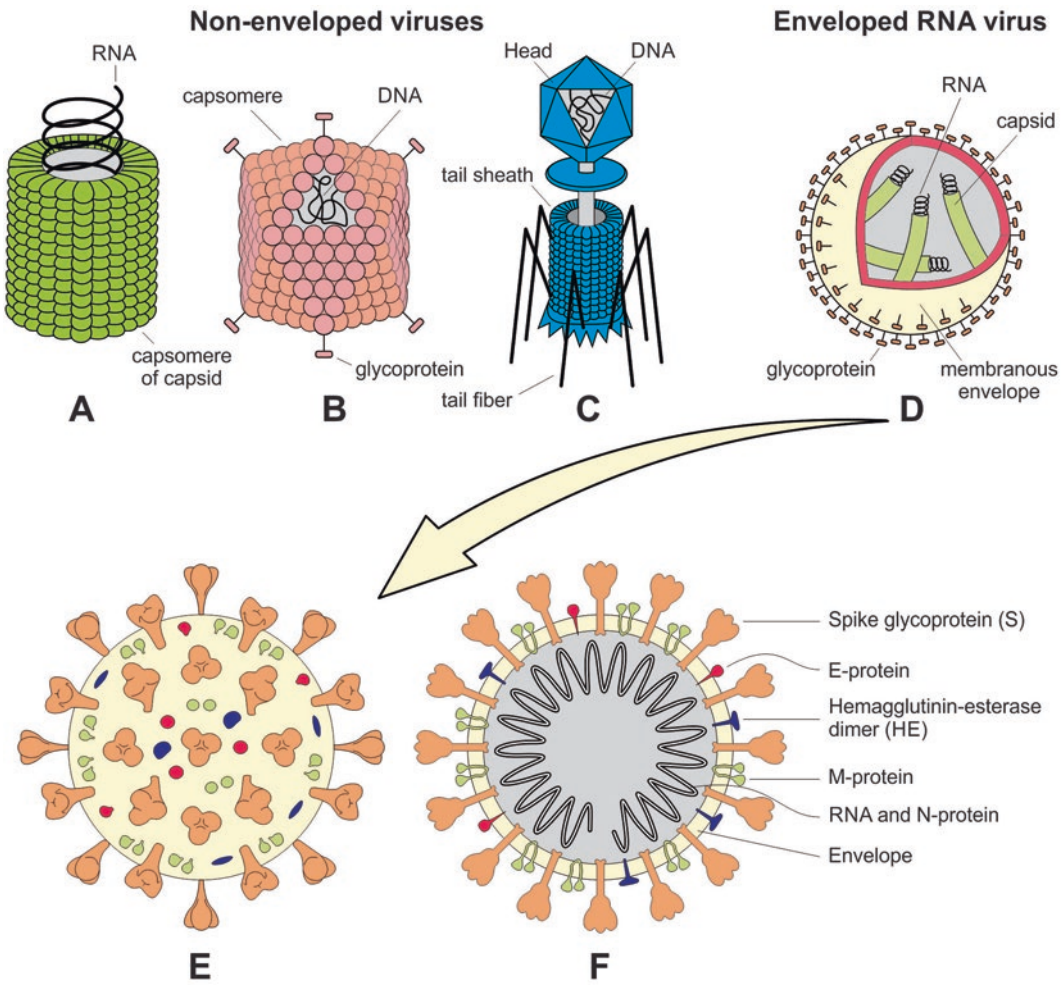


Fig. 15.1 Structure and shapes of different kinds of viruses. Viruses can be non-enveloped and enveloped. (**a**) A helical RNA virus; (**b**) an icosahedral DNA virus; (**c**) a complex DNA virus-like bacteriophage with a head and a tail; (**d**) an enveloped RNA virus; and (**e**, **f**) a novel coronavirus like COVID-19 and enveloped RNA virus with exterior and cross-sectional views, respectively

membrane. Some of these viruses also have proteins that stick out of the envelope or off the surface of the capsid. These proteins called spikes help the virus attach to host cells. Enveloped viruses include coronaviruses, human immunodeficiency virus (HIV), viruses that cause hepatitis B and C, herpes, Ebola, Zika, and dengue.

Viruses come in four standard shapes. Among them, helical, icosahedral, and complex viruses are non-enveloped, classified based on the architecture of their protein capsid, whereas the more derived one is enveloped.

'Helical' viruses have a capsid that forms a twisting helix around the nucleic acid (RNA) core (Fig. 15.1a). They are made up of a single protein subunit stacked around a central axis to form a helical structure. The tobacco mosaic virus (TMV) is an example of a helical virus.

'Icosahedral' viruses have a regular geometric shape of icosahedrons with 20 faces (Fig. 15.1b). Icosahedral capsid symmetry gives a virus a spherical shape under low magnification, but the protein subunits are arranged in a regular geometric pattern, like a geodesic dome. An icosahedral form is optimal for creating a sturdy structure from multiple copies of a single protein. Examples of icosahedral viruses are poliovirus, rhinovirus, and adenovirus.

'Complex' or head and tail viruses have separate patches of proteins, often forming unique structures or extensions on the virus (Fig. 15.1c). These viruses possess a capsid head, which is an elongated icosahedral, but the tail is helical and is surrounded by a protein coat. The tail may consist of a base plate with protein tail fibers. Viral protein subunits will self-assemble into a capsid, but the DNA of complex viruses also codes for proteins to help build the capsid. Many bacteriophage viruses are complex-shaped. They inject the cell with a viral genome capable of coding proteins to fight the cell's defenses. Another example of a complex virus is poxvirus, such as the variola virus, which causes smallpox.

'Enveloped' viruses can surround themselves in a portion of their host's cell membrane (Fig. 15.1d, e). The virus can use the host cell's outer membrane, internal membrane, or

endoplasmic reticulum. The membrane is studded with proteins coded for both the viral genome and the host genome. The influenza virus, hepatitis C, HIV, and the coronavirus are enveloped viruses.

15.2.1 Coronavirus (COVID-19)

Coronaviruses are a dreadful example of enveloped viruses that are pathogenic to mammals and birds. The pandemic of the coronavirus infectious disease at the time of writing was caused by an emergent strain of coronavirus, known as COVID-19 (severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)). In humans, these viruses cause severe respiratory syndrome that can range from mild to lethal. COVID-19 is unprecedented owing to the combination of its easy transmissibility and the extent to which it has disrupted the world. As of July 2022, nobody was able to predict how the pandemic would end. Many countries have entered the endemic stage of the COVID-19 outbreak; the virus is widespread, but is significantly less fatal than it was in 2020. More than seven million people died from the COVID-19 outbreak worldwide as of July 2022, of which the death toll in the United States surpassed one million. The World Health Organization (WHO), on the other hand, estimates that nearly 15 million people were killed either by coronavirus or by its impact on overwhelmed health systems during the first 2 years of the pandemic, more than double the current official death toll of more 7 million.

COVID-19 is an enveloped, positive-stranded RNA virus with a halo of club-shaped spikes of protein that the virus uses to invade human cells (Fig. 15.1e, f). It measures around 125 nanometers across and consists of 30 genes. COVID-19 has extraordinarily large single-stranded genomes—approximately 26,000–32,000 bases of RNA. We can compare this virus with the simplest and smallest present-day cells, the mycoplasmas. Mycoplasmas are small bacteria of the degenerate type that generally lead to a parasitic existence close to animals and plants. Some have a diameter of 300 nm and comprise about 500 genes, which can direct the synthesis of similar numbers of proteins. The viral envelope of COVID-19 consists of a lipid bilayer, in which membrane (M), envelope (E), and spike (S) structural proteins are anchored. A spike protein is the primary determinant of viral tropism and is responsible for receptor binding and membrane fusion. The virus usually appears spherical, as seen under an electron microscope, with a crown or 'corona' of spikes on the surface. Inside the envelope is the nucleocapsid formed from nucleocapsid (N) proteins bound to the single-stranded RNA genome that allows the virus to hijack human cells and turn them into virus factories.

The large, positive-sense RNA genome of COVID-19 is associated with the nucleocapsid protein, forming a continuous 'beads-on-a-string'-type conformation (Fig. 15.1e, f). The persistence of viruses is partially due to their ability to change rapidly and adapt to new situations. Coronaviruses can be transmitted to people from animals such as bats and are called zoonotic. They replicate their genomes using enzymes called RNA-dependent RNA polymerases (RdRps), which are prone to mutations. However, so far, genomic analysis has suggested that COVID-19 mutates slowly, reducing the chance that it will evolve to become more deadly such as the Delta variant that causes more infections and spreads more easily than the earlier forms.

The COVID-19 virus likely first emerged in November 2019 in Wuhan, the capital city of the Hubei province in China and a hub of auto and steel manufacturing, which was the epicenter of the current outbreak of COVID-19 [9]. The virus's first emergence in Wuhan remains a mystery: did it come from a military laboratory (the Wuhan Institute of Virology) or a wet market? The claim from various intelligence sources that COVID-19 may have leaked, accidentally or otherwise, from a military laboratory is gaining currency. Two possible hosts of this deadly virus appear to be the Chinese pangolin and the greater horseshoe bat. New studies in the *Journal of Science* reveal that the Wuhan Seafood Market was most likely the epicenter of the COVID-19 pandemic in late 2019.

The only tool we can use to fight viruses is vaccination. The Chinese scientist Zhang Youngzhen posted the mRNA sequence of COVID-19 on an open-access site, virological.org. With the virus genetic code in hand, scientists around the world began work on a vaccine. The mRNA structure provided critical information for developing the coronavirus vaccine. As of 2022, three effective coronavirus vaccines were available in the United States, one by Pfizer/BioNTech, the second by Moderna, and the third by Johnson & Johnson. Instead of the live virus, Pfizer and Moderna vaccines use synthetic COVID-19 mRNA, which instructs our cells to make the 'spike protein' that triggers an immune response inside our bodies. This surface protein of the virus causes COVID-19. At the end of the process, our bodies have learned how to protect against future infections. The viral mRNA we are receiving will decompose in a few days because RNA is inherently a fragile molecule (that, by the way, is why the vaccine needs to be kept frozen), and it will no longer exist in our body. Vaccinated people gain this protection without ever having to risk the severe consequences of contracting COVID-19. Both Moderna and Pfizer vaccines have a 95% efficacy rate, but need two shots, followed by two boosters. The single-shot COVID-19 vaccine by Johnson & Johnson (Janssen) could make a major difference in getting the pandemic under control. It does not require ultracold storage conditions, as vaccines made by Moderna and Pfizer do, and regular refrigerators can be used to store this vaccine. Instead of using mRNA, the Johnson & Johnson vaccine is a viral vector vaccine. It uses a disabled adenovirus to deliver the instructions, but it has a considerably lower efficacy rate, about 66% than the Pfizer/Moderna vaccine.

15.2.2 Viral Proteins and Enzymes

A viral protein has a dual function: it is both a product of a virus and a component. Viral proteins are grouped according to their function as structural proteins, nonstructural proteins, regulatory proteins, and accessory proteins. Most viral structural proteins are components of the capsid and the envelope of the virus. The genetic material of a virus is stored and protected within the capsid.

The proteins encoded in viral genomes belong to three major classes. First, most viruses encode enzymes required for replication of the genome and production of mRNA from it. RNA viruses must encode an RNA polymerase or replicase since cells do not usually replicate RNA. Second, viruses must encode proteins that are used in the assembly of progeny viruses. For simpler viruses, these may consist of only one or a few structural proteins that assemble with the genome to form the progeny virion. Third, the larger viruses encode proteins that interfere with the defense mechanisms of the host.

Viral polymerases play a major role in viral genome replication and transcription. Based on the genome types and specific types of viruses, a variety of hallmark genes encode enzymes: RNA-dependent RNA polymerases (RdRps), RNA-dependent DNA polymerases (RdDps), reverse transcriptase (RT), DNA-dependent RNA polymerases (DdRps), and DNA-dependent DNA polymerases (DdDps). Viruses would contribute these enzymes to the protocell's evolution alongside recurrent infection, leading to the emergence of DNA and its transcription and replication during the origin of life. Viruses not only exchanged genes with the host protocells but also contributed critical enzymes to protocells by horizontal transfer of digital and analog information systems, a key evolutionary driver of the origin of life.

Viruses constantly change through mutations, and new variants are expected to occur over time. Sometimes new variants emerge and disappear. At other times, they emerge and persist. Multiple variants of the COVID-19 virus have been detected in the United States and globally during this pandemic. In 2021, the WHO introduced the nomenclature using Greek alphabets to designate variants such as Alpha, Delta, and Omicron.

15.2.3 Nucleic Acids

Viral genomes exhibit extraordinary diversity in nucleic acid type, size, complexity, and the information transfer pathways that they follow. A virus has either RNA or DNA genes only, never both, and is called an RNA virus or a DNA virus, respectively. The genomes of viruses may be single-stranded RNA (ssRNA), double-stranded RNA (dsRNA), single-stranded DNA (ssDNA), or double-stranded DNA (dsDNA). The function of a virus is to replicate itself and increase its population. Viruses have evolved various strategies to replicate their genomes and produce the structural and catalytic proteins needed to form new viruses. Viral genomes are expressed from mRNAs that are the nucleic acid of positive sense ((+) sense) ssRNA viruses or transcripts from the negative sense ((−) sense) of dsDNA or from ds or ss DNA viruses.

Three broad classes of viruses can be recognized based on their genomes and information flow, which may have independent evolutionary origins. One class contains DNA as the genome, irrespective of whether single- or double-stranded, and direct DNA replicates the DNA genome → DNA copying. The second class of viruses contains RNA as their genome, and the RNA is replicated by natural RNA → RNA copying. The third class of viruses, retroviruses, encodes the reverse transcriptase (RT) enzyme, and these viruses have an RNA → DNA step in their life cycle. These kinds of viruses would play critical roles in the origin of DNA in protocells. All viruses are haploid, that is, they contain only one copy of the genomic nucleic acid. The exceptions are retroviruses, which are diploid and contain two identical copies of the single-stranded genomic RNA.

Despite the astounding variety in the nucleic acids present in virus particles, these nucleic acids must be replicated by standard Watson–Crick mechanisms. Every coding strand must be copied into a noncoding strand of opposite polarity, and vice versa. This uniformity of replication between viruses and cells helps viruses to exploit the ribosomes of the hosts during translation and replication.

15.3 Viruses and Evolution

Viruses are not self-sustaining and need to enter a host cell to complete their life cycle. Viruses carry out parasitic roles far more than cooperation in the evolutionary process. Therefore, we tend to regard viruses only as pathogens, thereby dismissing their crucial importance in the evolution of life. Viruses cause disease but are also useful tools for the study of molecular and cellular biology. Furthermore, the development of viruses as vectors for the expression of foreign genes has given them a new and expanded role in science and medicine, including their potential use in gene therapy.

The sustained war between pathogens and their hosts has long been recognized as a key driver of evolution. Viruses form a fascinating and unique component of the evolutionary system. They directly exchange genetic information with living organisms—that is, within the web of life itself. Therefore, they indeed have effects on evolution that are faster and more direct than those of external forces such as climate changes or plate tectonics that simply select among more slowly generated internal genetic variations. The wide

distributions of viruses, combined with their rapid rates of replication and mutations, make their hosts significant drivers of genetic innovation [6]. Viruses affect Earth's life, from bacteria to plants to human populations, often determining which will survive.

We argue that RNA viruses contained relicts of the protein/RNA world even before the emergence of the first cell. Viruses are the probable precursors of the first cells, but they have helped shape and build genomes of their hosts, including humans. The impact of viruses on life is dramatic. The symbiotic relationship between viruses and cells is not always restricted to parasitism but extends to a wide range of mutualism. Most known viruses are, in fact, persistent and inapparent, not pathogenic (toxic). Viruses, though, cause sickness and death; they are one of the most creative forces in the history of life. Part of this directly derives from the harm they cause: any life-forms that can defend themselves against one virus or another will tend to leave more descendants than the less fortunate beings. Viruses have, therefore, led life-forms to evolve an impressive array of antivirus defenses. Many viruses are beneficial to their hosts and perform essential functions in others. Viruses are significant drivers of evolutionary transitions [6, 7].

However, the more profound creativity of viruses comes from their donation of genes and enzymes to their hosts during infection. Viral genomes can be incorporated into the genetic material of the host. Viruses move genes across the three domains of life through 'horizontal gene transfer' (HGT). Such genetic exchanges accelerate evolution. Less appreciated is the exchange of polymerases or enzymes between viruses and host protocells through the 'horizontal enzyme transfer' (HET). If the HGT is a transfer of the digital information system, the HET is an example of transfer of the analog information system. Viral enzymes played critical roles in the origin of life. Viral polymerases play a central role in DNA replication and transcription. Several steps in the virus life cycle require the activity of a polymerase. Based on the genome type and specific needs of a particular virus, a variety of enzymes are contributed by viral hallmark genes encoding proteins. These are RNA-dependent RNA polymerases (RdRps), RNA-dependent DNA polymerases (RdDps), DNA-dependent RNA polymerases (DdRps), and DNA-dependent DNA polymerases (DdDps). Viral polymerases slowly transformed RNA viruses into DNA viruses during recurrent infections of RNA protocells. The evolutionary networks of primordial viruses, their recurrent infections of protocells, and their polymerases accelerated the origin of the first cells. Two capsid proteins that are most widely distributed among viruses are the jelly roll capsid (JRC) proteins and the superfamily three helicase (S3H) [5].

Viruses mediate essential processes in the biosphere. They play a critical role in the cycling of carbon in the biosphere during the lysis of bacterial cells inside which they reproduce, thus circulating carbon in the oceans. They can mediate the transfer of genetic information from one cell to another as they infect and reproduce inside cells. Their role in information transfer among prokaryotes by horizontal gene transfer (HGT) might have triggered the evolutionary processes for the origin of archaea, eukaryotes, and multicellular organisms.

The history of life is a story of the symbiotic coevolution of viruses and their cellular hosts. All cellular life harbors diverse genetic parasites, including transposons, plasmids, viruses, and other selfish elements. The parasite–host coevolution is a major aspect of the evolution of life [5, 8]. Coevolution is often described as an incessant arms race. The billion-year war between viruses and cells is the major source of evolutionary novelties. Viruses evolve, the host adapts, proteins change, and viruses evade them. It never ends. Many innovations first selected in the viral world might have been transferred to cells due to the continuous flow of viral genes into cellular genomes. The war has driven a dramatic diversification of viruses and of the host defense system. Viruses have a remarkable capacity to invade, replicate, and evolve within living cells. In response, cells developed an array of defense systems. Viruses and protocells have been intertwined since the protein/RNA world. Viruses may have evolved alongside the very first 'living' cells. Viral replication within a living cell always produces changes in the host cell, sometimes resulting in cell death and sometimes slowly killing the infected cells.

The creative role of viruses in the origin and evolution of life has been known for a century. Viruses are truly nature's genomics laboratory, and they help accelerate the host's evolution in the fast lane. Felix D'Herelle, the discoverer of bacteriophages and one of the founders of virology, proposed as early as 1922 that phages or bacteriophages might be the evolutionary precursors of cells [10]. Similarly, in his 1928 classic *The Origin of Life*, J.B.S. Haldane suggested an early 'viral' stage of evolution as an integral part of the proposed scenario for the emergence of life from the prebiotic soup [11]. In our present discussion on the origin of life, we have followed and elaborated on Haldane's insight.

Viruses are potentially aggressive, selfish elements that develop a parasitic partnership with protocells; this association has powerful evolutionary outcome. Viruses are adept at transferring genetic information between themselves and hosts. They contributed several viral enzymes to the host protocells, including RNA-dependent RNA polymerases, DNA-dependent RNA polymerases, and DNA-depended DNA polymerases during RNA–DNA transition. RNA viruses might have coexisted with early protocells that still had RNA genomes. Because viruses have such ancient roots, they possess a remarkable range of biochemical tricks [12]. Viruses must establish an intimate relationship with their hosts and vectors to infect, replicate, and disseminate; hence,

they can be considered symbionts—intimate partners—of their hosts.

The great billion-year war between viruses and cells is the primary source of evolutionary novelties [3]. Reverse transcription paved the way for generating DNA. DNA is generated from RNA in retroviruses, cancer cells, and HIV. Viruses donated DNA and replicating genes to protocells [7, 12]. They might have played a central role in the emergence of eukaryotes and their nuclei [13, 14]; they might have been the cause of the partitioning of biological organisms into three domains of life—bacteria, archaea, and eukaryotes—by horizontal gene transfer [15, 16]. Their role in information transfer between extant prokaryotes by horizontal gene transfer complicates efforts to build evolutionary trees depicting early life on Earth and to unravel the origin of specific metabolic pathways.

Many viruses have their own ancient evolutionary history. Viruses possess genes, replicate, evolve, and adapt to specific hosts, biotic habitats, and ecological niches. From prebiotic protocells to unicellular life to human populations, viruses affect life's outcomes and provide an ever-changing shape to the fitness landscape, often determining which organisms will survive [5]. Since the beginnings of life, viruses have been the major drivers of macroevolution in all branches of life by horizontal gene transfer. They comprise the principal source of novel genes in the biosphere [5, 17].

Unlike most viruses, which infect, replicate, and leave their hosts, retroviruses integrate their genetic material into the cells they infect. If this happens to be a germline, then the viral genome (or its relict) can be permanently maintained in every subsequent generation. The retrovirus and the host become one. Retroviruses have invaded the genome of human ancestors over the course of a million years, yet these viruses are generally inactivated during evolution, with only remnants of these infectious sequences remaining in the human genome. Each of us carries almost a 100,000 fragments of endogenous retrovirus (ERV) DNA in our genome. Human endogenous retroviruses (HERVs) make up 8% of human DNA, with HERV-K colonizing the genomes of ancient primates as early as 55 million years ago and embedding their viral genes into our genomes [18]. To put this figure into perspective, consider that 20,000 protein-coding genes in the human genome make up only 1.2% of our DNA. HERV-K may protect human embryos from other viruses. Some of these genes now play crucial roles in the early stages of the developing embryo and the placenta that surrounds the fetus [19]. About 100 million years ago, a symbiotic retrovirus enabled it to evolve a placenta over many generations in mammals in the shadow of the last dinosaurs. In order to let a fetus mature inside a mother's uterus, an animal needed a way to provide oxygen and nutrients while removing waste and keeping both blood supplies separate. Scientists are continuing to untangle how viruses have become part of us. A virus–host interaction is a significant evolutionary force and played a crucial role in the origin and evolution of life.

Selfishness is pervasive and manifests at all scales of biology, from societies to individuals and to genetic elements within a genome. A virus is no exception. The relentless struggle to seek evolutionary advantages drives perpetual cycles of adaptation and counter-adaptation, commonly referred to as Red Queen interactions. For example, many RNA viruses mutate rapidly to overcome the varied challenges imposed by a cell's immune systems. However, a higher mutation rate may cause a progeny virus to carry a deleterious mutant, and the viral populations have to adapt continually. According to the 'Red Queen hypothesis,' both the parasite and host are perpetually struggling to maintain a constant fitness level. This long-term evolutionary pressure gave rise to surprising consequences for the entire tree of life [20].

Bacteriophages are not forgotten viruses. The emergence of pathogenic bacteria, resistant to most, if not all, currently available antimicrobial agents, has become a critical problem in modern medicine. Phage therapy, the use of bacteriophage viruses to treat bacterial infections, has existed for more than a hundred years, predating antibiotics. Bacteriophage is a potential solution in the fight against the rise of antimicrobial resistance. It is widely being reconsidered as an alternative to antibiotics.

15.4 The Origin of Viruses

The origin of viruses is shrouded in mystery, but recent advances in genomics have shed light on their ancient ancestry. Viruses have never been detected in fossils, probably because they are too small and fragile for fossilization. Therefore, the evolutionary history of viruses is difficult to reconstruct. Tracing the origins of viruses is difficult because of the tricks they use to make copies of themselves within the cells they have invaded. Some viruses even can stitch their genes into those cells they infect by horizontal gene transfer, which means studying their ancestry from molecular phylogeny requires untangling it from the history of their hosts and other organisms. Despite these difficulties, scientists have been able to piece together the likely origin of viruses from their hallmark genes and enzymes.

No one knows when and where viruses originated, but many biologists suggest that they are closely related to plasmids and transposable elements such as 'jumping genes' [6]. Viruses, plasmids, and transposable elements are all acellular, mobile genetic elements that replicate with the aid of a host cell, called replicons [5]. Simple viruses are indistinguishable from plasmids, except for one feature: they have a protein coat or a membrane-like envelope.

For many years, the central debating point in discussions of the origin of viruses has been whether they are ancient, first appearing before the last universal common ancestor (LUCA), or whether they have evolved more recently, such that their ancestry lies with genes that 'escaped' from the genomes of their cellular host organisms and subsequently evolved through independent replication. Historically, three main hypotheses for the origins of viruses have been proposed [21]:

1. The 'virus-first hypothesis' or the primordial virus world claims that the ancestors of viruses are direct descendants of the first replicons (pieces of genetic information capable of self-replication) that existed alongside the protocells (Fig. 15.2a). They arose from complex molecules of RNAs and proteins in the prebiotic world, infected the protocells, and contributed to the rise of DNA and the first cells; D'Herelle first proposed the hypothesis about a century ago, asserting that viruses are the ancestors of cells [10].
2. The 'regression' or reduction hypothesis asserts that viruses are remnants of small cellular organisms that parasitized larger cells (Fig. 15.2b).
3. The 'escaped gene' or progressive hypothesis states that viruses evolved from multiple, independent occasions from different organisms from host genes, which became independent of their host cells (Fig. 15.2c).

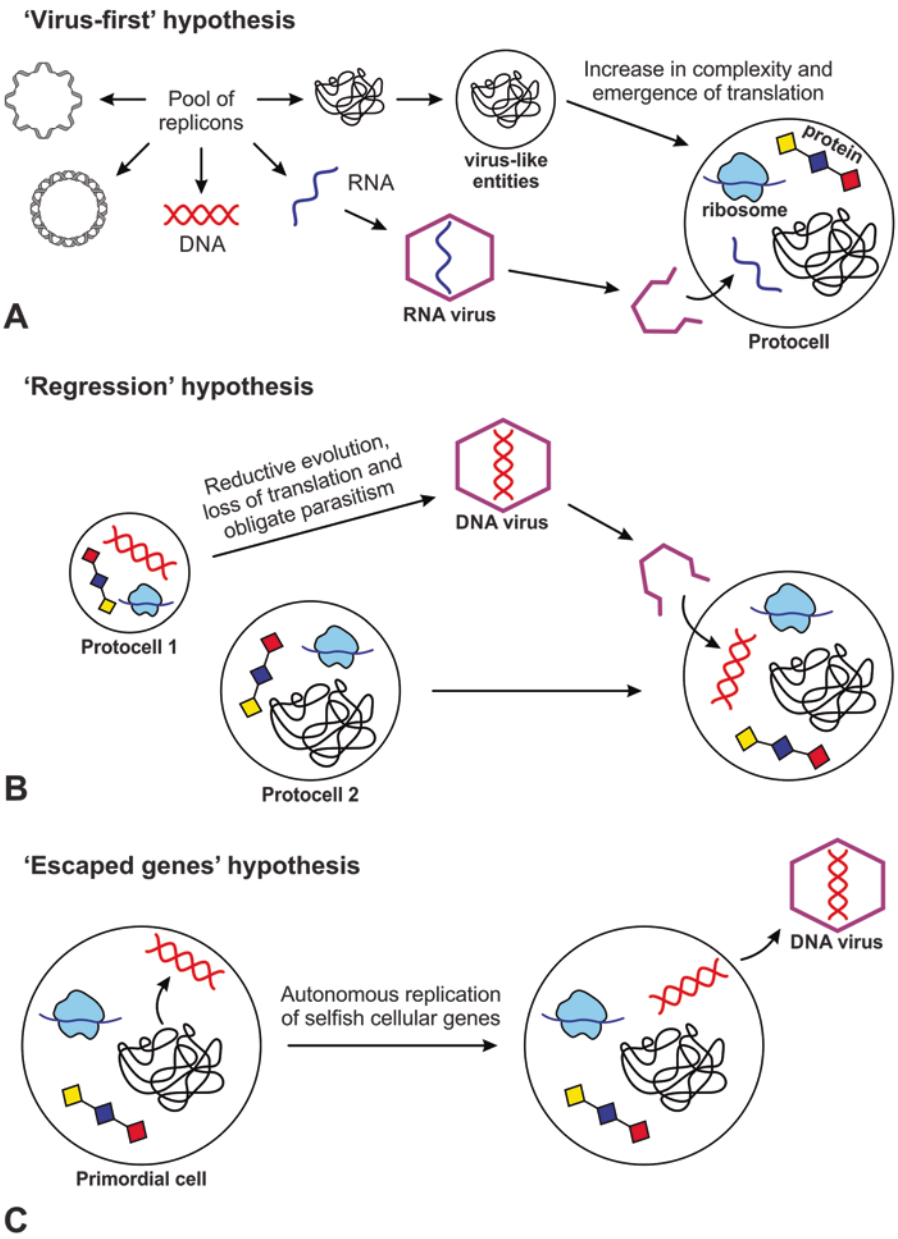


Fig. 15.2 Three major scenarios for the origin of viruses; the arrows show the direction of evolutionary changes. (**a**) According to the 'virus-first hypothesis,' viruses evolved from early replicative elements that preceded the first cellular life-forms; in the protein/RNA world, the first RNA virus evolved from hybrid molecules of RNA and proteins and began to infect protocells, hijacking their ribosomes for replication, and coevolved with their hosts. (**b**) According to the 'regression' hypothesis, viruses emerged through degeneration of cells that then assumed a parasitic lifestyle. (**c**) Finally, according to the 'escaped gene' hypothesis, viruses derived from DNA/RNA that escaped from a primordial cell and acquired the ability for 'selfish' replication and spread. (modified from [23])

Over the years, all three scenarios have been revised, elaborated, temporarily discarded, and revisited. Among these, the escaped gene hypothesis has traditionally been popular from the uniformitarianism principle (the present is the key to the past) because viruses are parasitic on cells now. It has been argued that this must have been the case in the geological past. However, in recent times, the virus-first hypothesis has been gaining momentum.

While the geological record cannot offer any clue of when and how viruses originated, genetics provides increasingly strong support for viruses' ancient primordial origin (Fig. 15.2a). Viruses are believed to be ancient, to have evolved in tandem with the first life-forms. First, most viruses do not encode genes for ribosomal proteins or furnish genetic evidence of relics of such genes. Second, the same vast majority of viruses do not contain genetic evidence of ever having encoded enzymes involved in energy metabolism. Third, viral capsid proteins typically do not have apparent homologs among contemporary cellular proteins. Several genes coding for key proteins involved in viral infection and major capsid proteins of icosahedral virions are shared by many groups of viruses but are missing in cellular life-forms. Therefore, most viral proteins have no cellular homologs or have only distantly related ones. All combined evidence argues that viruses did not evolve from free-living cells but arose independently in the prebiotic world before the first cells. The existence of hallmark genes seems to falsify both the cell degeneration and the escaped gene concepts of viral infection. This genetic disparity between viruses and living cells strongly supports the model of an ancient virus world, a flow of virus-specific genes that continued uninterrupted from the precellular stage of life's evolution to this day [5–8, 21–28].

Viruses effectively use all possible genome replication strategies and expression, with different nucleic acid forms (ssRNA, dsRNA, ssDNA, or dsDNA) as the genome that is incorporated into virions. This diversity of genomic strategies in viruses contrasts with the uniformity observed in cellular organisms and seems to be best compatible with the possibility that the virus world descended directly from a precellular stage of evolution [27, 28]. This virus-first hypothesis is supported by the strong inverse relationship between genome size mutation rate across all replications systems, such that pre-LUCA genomes were probably both small and highly error-prone and hence RNA-virus-like [21]. I endorse the virus-first hypothesis, like many of my predecessors, based on growing evidence in the protein/RNA world.

The early viruses have two hybrid components: the outer protein capsid shell (analog) encasing the RNA core (digital). These two components provide a timeline for the virus's origin in the protein/RNA world during the origin of life. Viruses removed the cumbersome translation machine from the protein/RNA protocell when inert. During infection of the protocells, they became alive and coerced the translation machine of the protocells for their reproduction. Viruses have invented a minimalist design for survival and have coevolved with protocells and cells for the last four billion years.

15.5 The Prebiotic Origin of mRNA Viruses

According to the virus-first hypothesis, the ancestors of viruses emerged early in Earth's history, from the primordial soup containing different species of RNAs and other replicons, amino acids, lipids, and proteins that formed in hydrothermal vent environments [5–7, 21–28]. These molecules also led to the evolution of protocells and ancestral viruses in parallel. Early viroids could have simply existed as free-floating pieces of nucleic acids. Ribozymes are closely related to viroids, which are virus-like elements with non-coding hairpin loop-structured circular RNA, sometimes with catalytic activity and without a protein coat that is typical of viruses. The earliest pieces of genetic material were likely short pieces of RNA with relatively few genes, with a protective coating of proteins that parasitized other floating protocells to make copies of themselves. These primordial viruses swapped genes with protocells containing RNA and translation machinery. Over time, the parasitic genetic elements, these primordial viruses, remained unable to replicate on their own and evolved into modern-day viruses that free-load on their cellular hosts. They developed a simple way of creating new viruses that required only a minimal investment of molecular machinery. All they needed to do was get a copy of a viral mRNA into a cell. They have preferred this minimalist-design lifestyle for the last four billion years. The variety of ways in which viruses store their genomic information (single- or double-stranded RNA or DNA), in contrast to cellular organisms, lends credence to the idea that viruses emerged before the appearance of the last universal common ancestor (LUCA).

The flow of virus-specific genes has gone uninterrupted from the precellular stage of life's evolution in the ancient viral world to the present. In the primordial genetic pool, RNA viruses would evolve first, followed by retroid elements and DNA viruses [3]. In our view, three major classes of viruses originated in the prebiotic world:

- Positive-stranded mRNA viruses in the primordial protein/RNA world
- Retroid viruses in the RNA–DNA transition world
- DNA viruses in the DNA world

These ancient viruses emerged in a hydrothermal vent environment in which the mixing and matching of diverse

genetic elements were more extensive than in any modern biological community. Phylogenetic analyses have shown that the RNA polymerase, DNA polymerase, and DNA helicase transcribing and replicating DNA in modern cells were recruited from the viral world [12, 19, 27].

The idea that viruses are extremely ancient and have coevolved with protocells has recently led to several hypotheses stating that viruses have played a significant role in several critical evolutionary transitions [2–5]. Viruses ubiquitously infect all members of the three cellular domains of life, suggesting that protocells with RNA genomes were already the victims of a viral attack. Moreover, viruses infecting cells from the three domains of life—bacteria, archaea, and eukaryotes—suggest that they emerged very early in the prebiotic world, before the first cells [3, 21, 22].

The notion of viral antiquity seems more natural to accept for mRNA viruses in the hydrothermal vent environment. High temperatures in the vent environment favored the high diversity of virus-like particles [6]. The only organisms with RNA-coded genomes today are RNA-based viruses, which may provide helpful insights into the protein/RNA world. The virus world retained a distinct flow of genes from the repeated infection of protocells containing RNAs, ribosomes, and proteins. Viruses have maintained their identities and unique parasitic lifestyle ever since, notwithstanding the transfer of many genes between viral and cellular genomes. Several genes central to viral replication and structures are shared by all viruses but are absent from cellular genomes. In this scenario, the principal lineages of viruses emerged from the primordial pool of primitive genetic elements with a distinct suite of viral genes that retained their identity throughout their entire life history. Viruses enhanced gene mixing in the prebiotic world by infecting protocells containing RNAs and protein molecules and by their subsequent endogenization. Therefore, the viral evolution in the prebiotic world is closely intertwined with the origin and early evolution of cells [24, 25, 28].

15.6 The Beginning of the Virus World

Viruses are mobile genetic elements (MGEs). The principal lineages of viruses and related selfish agents emerged from the primordial genetic pool of primitive genetic elements in the hydrothermal vent environment, the ancestors of both cellular and genetic elements. The primordial gene pool was a crucible for the major virus lineages, where mixing and matching diverse genetic elements was extensive. Viruses reflect their origin from capsidless selfish replicons, such as plasmids, transposons, and viroids. In this scenario, viruses are direct descendants of primordial genetic elements [3]. These selfish replicons, called viral hallmark genes, encoded capsid proteins with key roles in genome replication, expression, and encapsidation. A wide variety of viruses share these selfish replicons. However, they are missing from cellular genomes, suggesting a flow of virus-specific genes that has gone uninterrupted from the protein/RNA world of biogenesis to the present day. These viral genes are genuine viral hallmarks and can originate either through the modification of existing genes or de novo. Eventually, diverse protein-coding RNA elements would develop a capsid coat, giving rise to the first viruses [3, 25]. The protein/RNA world is the lower limit for the origin of virus.

15.6.1 Pre-virus

We concur with the ancient virus world hypothesis [5–10, 22–29]. In our model, ancestral viral particles (called the 'pre-virus') could have emerged only in the protein/mRNA world, when selfish viral genes modified from existing mRNA molecules de novo overprinted by a new open reading frame [27]. These viral genes were dispersed in hydrothermal vents. Since mRNA molecules were fragile, some of the viral genes were encased by proteins randomly for stability and durability to form the pre-virus (Fig. 15.3a). Some of the viral genes were dispersed in the prebiotic soup, making it an ideal nature's genomics laboratory. Different genetic innovations took place in the genetic pool of the hydrothermal vent that mixed, matched, and evolved new, increasingly complex viral gene ensembles.

15.6.2 mRNA Virus

Perhaps some of the ancestral pre-mRNA viruses were accidentally ingested by the protein/RNA protocells by infolding their membranes while searching for food in the vent environment. This ingestion of foreign material by phagocytic protocells might have been the beginning of endocytosis. Inside the protocell, the viral genes began to exploit the protocell's translation machinery to make custom-made capsid proteins. The two important capsid proteins are the jelly roll capsid (JRC) proteins and the superfamily 3 helicase (S3H). The former is more widespread than the latter [25]. We speculate that the primordial mRNA gene-encoding capsid protein, JRC, was present in the vent environment (Fig. 15.3a, b). New JRC protein genes can originate either through the modification of existing genes or de novo. Viruses contain many de novo genes, namely, those in which an existing gene has been 'overprinted' by a new open reading frame; mutations of the mRNA gene led to the expression of a second reading frame, overlapping the first. Overlapping genes are widespread in viral genomes [31].

Once some mRNA molecules developed the capsid coat, the first mRNA viruses originated. The capsid affords

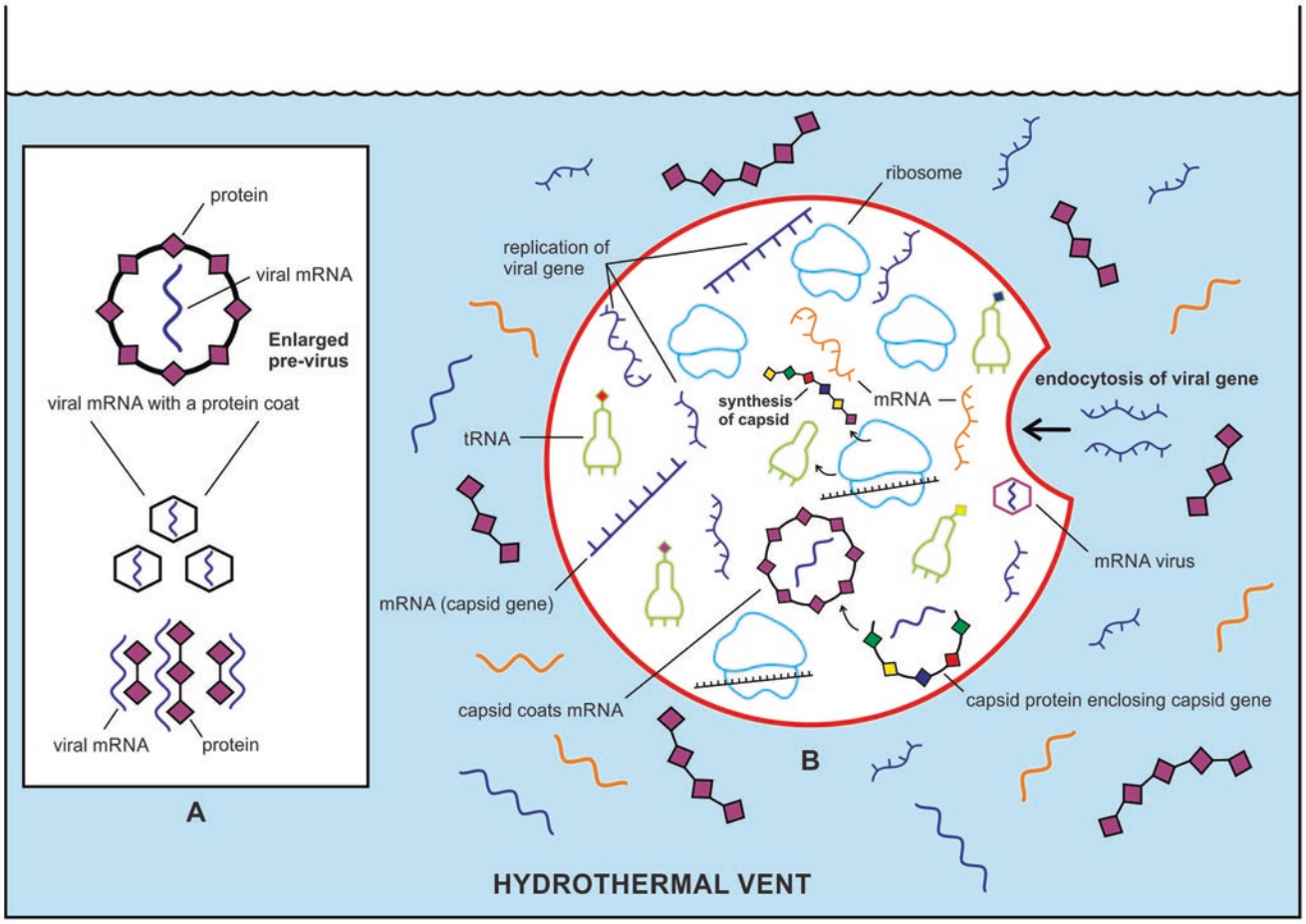


Fig. 15.3 The ancient viral world began in the prebiotic soup in the hydrothermal crater vent environment, which was an ideal nature's genomics laboratory. (**a**) In our model, selfish viral genes modified from mRNA molecules appeared de novo by overprinting in the prebiotic soup; a protein shell encased these capsid genes for stability and durability. However, these proteins were random, readily available in the prebiotic soup, and were not encoded by viral genes. This initial stage of the viral structure that afforded protection to fragile mRNA is called the pre-virus. (**b**) In the next step of evolution, some viral genes entered the protocells by endocytosis, utilizing their ribosomes for the synthesis of capsid proteins. Once the capsid protein began to coat the viral gene, the first mRNA virus appeared

protection to the viral gene and allows viral genes to access appropriate host protocells. As more and more mRNA viruses were created, they exerted osmotic pressure on the phospholipid membrane, causing a burst of the protocells (Fig. 15.3b). Slowly, these newly released viruses learned by trial and error how to infect protocells and swap genes with them prior to inventing their translation machinery. One of the key features of ancestral mRNA viruses was their reliance on protein/RNA protocells for replication and propagation. Specifically, they depended on host cells for a translation machinery for synthesis of viral proteins from viral mRNAs and solved this problem by infection. This innovative shortcut strategy for gene swapping for virus replication inside protocells has worked efficiently, hijacking the host's translation machinery. Viruses have preferred this parasitic existence from the beginning of biosynthesis and have continued to proliferate throughout the geological ages.

15.7 The Replication Cycle of mRNA Viruses: Digital Renaissance

The horizontal transfer of viral genes to the host protocells and using their ribosomes for making capsids brought about a digital renaissance and a significant milestone for abiogenesis. The emergence of the ancestral virus with a capsid-coding sequence of proteins was a significant evolutionary step and was mediated by ribosome-coding protocells. Here, we propose a model for the parallel evolution of protocells and emerging viruses. As more and more proteins were synthesized, various kinds of protocells dominated the hydrothermal vent environment. Some of these protocells were densely packed with diverse populations of genetic elements, including self-replicating mRNAs, multiple protein-coding mRNAs, and translation machinery in their cytoplasms (Fig. 15.3).

A typical viral genome encompasses two core modules: genes encoding proteins required for genome replication and

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

Viruses can reproduce only within host protocells, exploiting their ribosomes. The parental virus (virion) produces numerous progenies, usually genetically and structurally identical to the parent virus. The new generation of viruses began to infect other protocells with an ensemble of translation machinery for their replication. In this way, viral mRNAs could be translated into capsid proteins inside the host protocells. They soon became obligate intracellular parasites, that is, they developed and reproduced only within the hosts' protocells. They hijacked the host's machinery to manufacture hundreds of copies of themselves. The reproductive successes of viruses made them archenemies of protocells. Viruses became the capsid-coding particles that began to coevolve with ribosome-encoding protocells. Because of their high mutation rates, viruses were evolution accelerators.

These primordial viruses were single-stranded mRNA viruses that had two components that could encode two different functions: (1) a replicon allowing genome multiplication and (2) a capsid not only to protect the viral genome in the extracellular space but also involved in the entrance and exit mechanisms of virions in and out of protocells [2]. Like living counterparts, the genome contains relatively few genes, usually between 3 and 10, and a critical viral enzyme such as an RNA-dependent RNA polymerase (RdRp) or an RNA replicase. This viral protein enzyme synthesizes mRNA from the mRNA template. An RdRp is an essential enzyme in the genomes of all RNA-containing viruses before the DNA stage. mRNA viruses would donate an *RdRp* gene to the host protein/RNA protocell by horizontal enzyme transfer (HET), thus facilitating replication of its mRNA. This mechanism could allow the emergence of the first RNA-based life, as discussed in the next section.

Perhaps, primordial virions initially did not kill or lyse their host protocells but utilized their translation machinery to reproduce genomes. Some of the likely stages of the cycle of infection were as follows (Fig. 15.4b):

1. *Attachment*: Capsids on the surface of the mRNA virus attach to the surface of the host protocell (Fig. 15.4b, cycle 1).
2. *Entry via endocytosis*: The virus enters the interior of the host protocell through the process of endocytosis.
3. *Uncoating of the capsid*: Inside the protocell, the viral genome emerged from the protein capsid; the capsid, in turn, destroyed the host mRNA so that viral mRNA occupied its place (Fig. 15.4b, cycle 2, inset).
4. *Copying mRNA and protein synthesis*: RdRp enzymes copy the viral genome. Energy and ribosomes from the host protocell are used to build viral proteins (Fig. 15.4b, cycle 2, and cycle 3).
5. *Assembly of viral progeny*: The viral particles assemble by encapsidation to form progeny virions (Fig. 15.4b, cycle 4).
6. *Release via exocytosis*: The virus exits the host protocell by exocytosis (Fig. 15.4b, cycle 4, inset).

The progeny virions began to infect other protocells to start the next cycle of infection (Fig. 15.4b, cycle 5). The life cycle of most viruses is designed to maximize the production of progeny virus particles. Often, the burden of producing a large number of virus particles causes the infected cells to die, resulting in the lysis of the host cell. In the early stages, primordial viruses probably established a long-term association with the protocell. The protocell released a steady stream of viral particles over an extended period, benefiting both host and parasite in symbiosis. These ancient RNA viruses had a high mutation rate and underwent an evolution and natural selection, just like cellular life, and most of them evolved rapidly. When two viruses infected a protocell simultaneously, they might have swapped genetic material to make new 'mixed' viruses with unique properties. The viral infection of the protocell is a prelude to a modern bacteriophage that infects and replicates within bacteria and archaea.

This stage begins with mRNA viruses and their spread in the vent environment. Today, mRNA viruses amount to a significant fraction of known viruses, including many pathogens, such as the hepatitis C virus, West Nile Virus, dengue virus, and SARS and Middle East respiratory syndrome (MERS) coronaviruses. They also include less clinically severe pathogens, such as rhinoviruses, which cause the common cold [30].

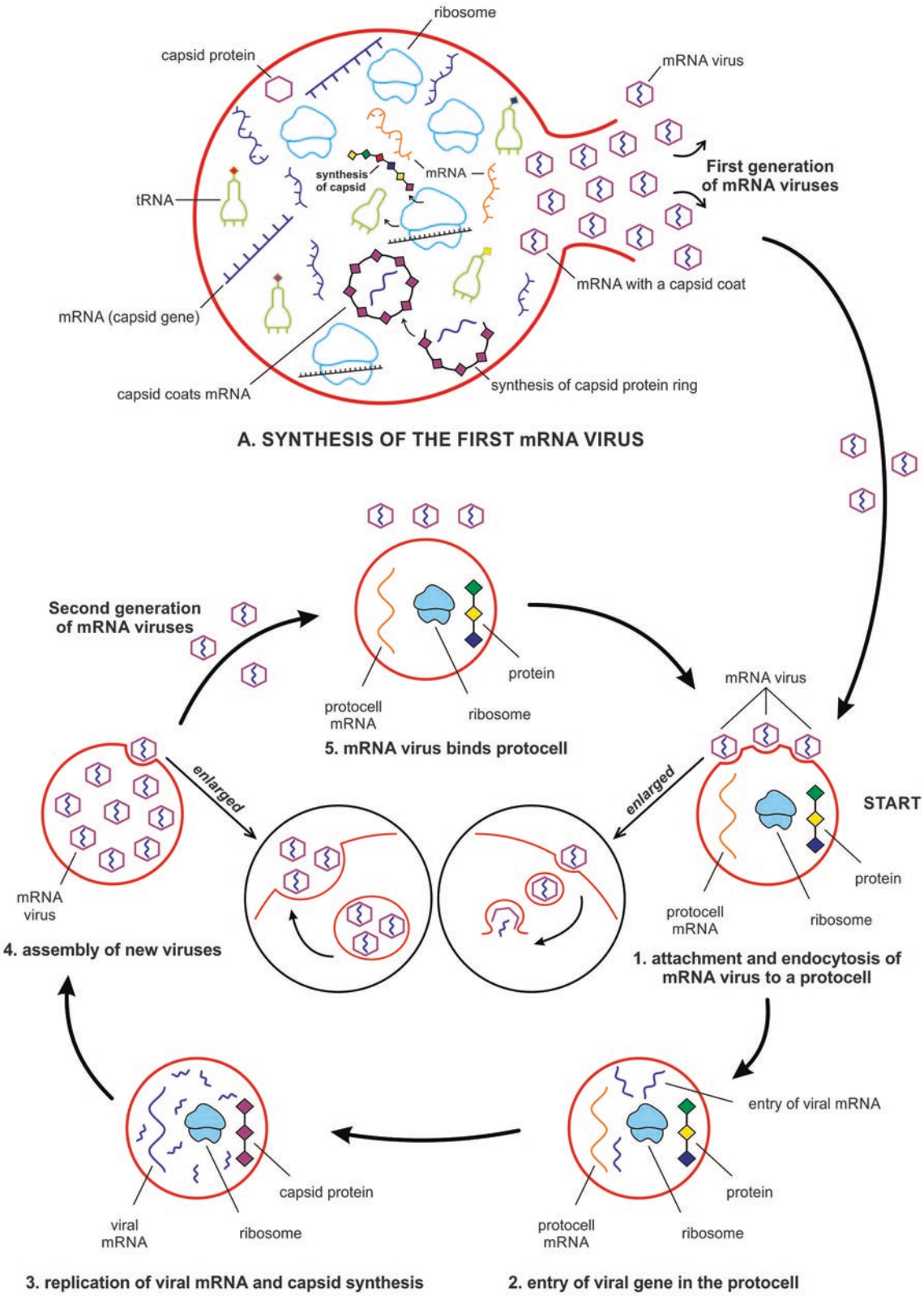


Fig. 15.4 The origin of mRNA viruses and the replication cycle. (**a**) Synthesis of the first generation of mRNA viruses inside protocells hijacking hosts' translation machinery. Viral *mRNA* genes produced capsid proteins that began to wrap viral genes. As more and more mRNA viruses were synthesized inside protocells, they exerted osmotic pressure on the plasma membrane, causing a burst of protocells, releasing the first batch of mRNA viruses for infection of protocells. (**b**) Some likely stages of the replication cycle of mRNA viruses includes attachment of a protocell and its entry into the protocell via endocytosis (cycle 1), uncoating of the capsid shell (cycle 2), mRNA copying and protein synthesis (cycles 2 and 3, respectively), and self-assembly of viral progeny and release via exocytosis (cycle 4). Newly released mRNA viruses began to infect protocells with ribosomes (cycle 5) to start the next cycle of infection

15.8 Primordial Protocellular Life

The proponents of the 'RNA world' scenario propose that RNA protocells began replicating their genome by the RNA polymerase ribozyme [32], but how RNA is replicated remains unclear. At that time, no enzyme was available. There are no naturally occurring ribozymes that can catalyze and replicate an RNA molecule. Despite many in vitro attempts, developing a model for nonenzymatic replication of RNA genomes by ribozymes is, at best, in its infancy [33]. It is unlikely that RNA protocells could divide by replicating an RNA strand to make viable early life. The emerging life could only form after the development of the genetic code, protein synthesis, and the donation of a unique viral RNA replicase to a protein/mRNA protocell.

RNA genomes dominate the world of viruses. Their success results from the possibility of accommodating rapid changes via mutations. In the protein/mRNA world, the success of mRNA viruses as pathogens depends on their ability to replicate within host cells. An RNA-dependent RNA polymerase (RdRp) is one of the most versatile enzymes of mRNA that is responsible for replicating the genome. As discussed earlier, mRNA viruses encode two different components, a replicon such as an RdRp, which allows the genome to multiply, and a capsid, which not only offers protection to the viral genome in a shell but also allows the entry and exit mechanisms of virions in and out of the host cells, respectively. Following the viral entry into the protein/mRNA protocell, the viral genes are released by HGT to swap genes of the protocells and are then translated to produce viral RdRps and capsids. In ssRNA viruses, RdRps use the positive genomic strand as a template to create a new negative-strand copy that is subsequently used as a template to create large numbers of positive-strand viral RNA genomes. RdRp is one of the most common RNA replicase that catalyzes the replication of RNA from an RNA template. However, RdRp was not available in the RNA world. During protein/mRNA protocell infection by an mRNA virus, the host cell provided a place for viral replication. In exchange, the mRNA virus might have contributed the RdRp enzyme to the protein/RNA protocell, which might have resulted in a diverse population of protocell types with high replication errors. RdRp is an essential enzyme encoded in the genomes of most mRNA-containing viruses with no DNA stage. During infection, mRNA viruses not only swapped genes of the host cells for reproduction but also contributed their replicating enzyme RdRp for gene duplication of the host cells by horizontal enzyme transfer (HET). Two systems, the gene of the protocell and the enzyme of the virus, melded together, which aided the replication. RdRp facilitated the reproduction of the host cells to produce identical daughter cells. One of the key features of viruses is their reliance on living cells for replication and propagation. An infection of a host cell and viral propagation is dependent on the translation of viral proteins as well as genome replication with simplicity and economy. The more daughter protocells of host parent protocells are created by the viral enzyme RdRp, the more host protocells would be available for viral infection. Thus, a viral infection provided an early glimpse of primordial protein/mRNA protocellular life as an example of mutual help of symbiosis (Fig. 15.5). From a virus's perspective, the more protocells that are available for infection, the more production and survival of its kind, a mechanism for its reproductive success. RNA viruses are said to replicate at the edge of 'error catastrophe' [34]. However, protocell genomes were more stable than were viral genomes. Non-catastrophic levels of genome instabilities in the protocells are instrumental for accumulating beneficial variants to meet the challenges of ever-changing environments.

In the peptide/RNA world, we have discussed the fission mechanism of protocells, where daughter cells might inherit an unequal amount of cytoplasmic content (see Fig. 9.4). This problem of unequal genetic transfer from parent to daughter protocells could be averted if the protocell contained all the mRNAs in a single gene that could be duplicated shortly before protocell division. During protocell division, each daughter protocell takes one of the two genes. The protein/mRNA protocell invented symmetrical fission, so the daughter protocells become genetically identical to their parent protocell, a prelude to the modern bacterial cell.

However, these protein/mRNA protocells with a single-gene level can encode a single enzyme. Since, after translation, the mRNA gene would be recycled, permanent storage of genetic information would be needed. When there was no DNA, RNA served both as a replicable repository of genetic information and as an agent in expressing that information. However, when the gene is expressed, the protocell has no function. At best, the protein/mRNA protocell represented an early stage in the development of life.

The advent of the protein/mRNA protocell and its subsequent viral infection introduced RNA replication, the essence of genetic continuity, and, with it, variation, competition, and selection. Life depends on the propagation of heritable information across successive generations. In all known life, heritable information is maintained within the sequence of DNA genomes copied by polymerase enzymes. In the protein/RNA world, this task would have been performed by the RdRp enzyme. If the protocell could divide symmetrically, then it would have something like a mini organism, the primordial life. Moreover, protein/mRNA protocells, with the development of a translation machine and the genetic code, could create other proteins essential for the survival of the protocells. The RNA theory exclusively focuses on replication at the expense of metabolism. Life is more than raw reproduc-

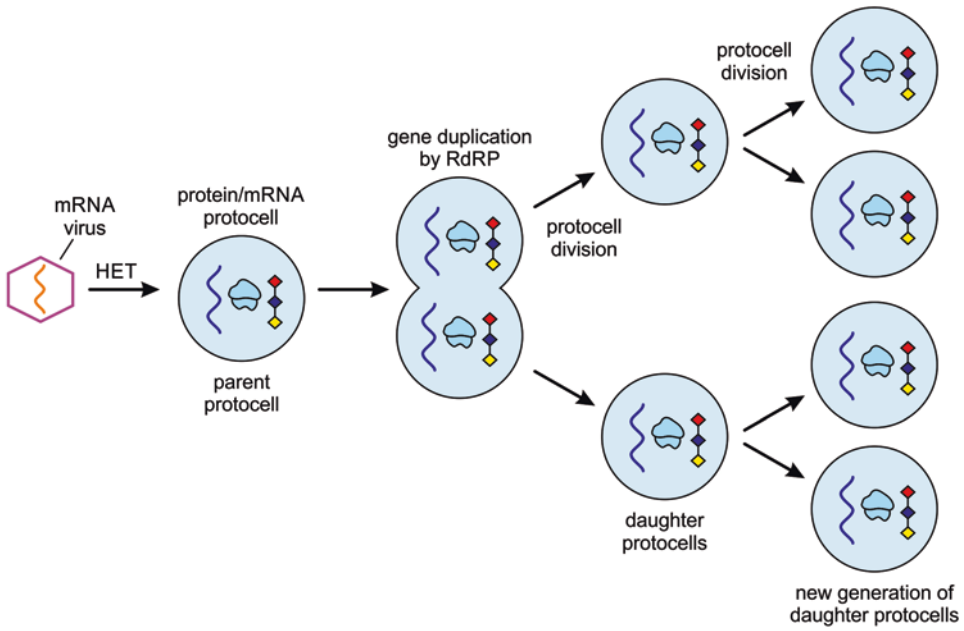


Fig. 15.5 The origin of primordial protocellular life during the horizontal enzyme transfer (HET) of a viral RdRp to a protein/mRNA protocell. RdRp facilitated the gene duplication and division of the protein/mRNA protocell. During protocell division, components of the three information systems such as digital (DNA), hybrid (translation machine), and analog (protocell membrane and cytoplasm) were transferred vertically from parent to daughter cells. Because of the high mutation of RNA replication, protocells would show genetic variation and be subjected to Darwinian selection in the protocell population. Once the protocell division was perfected, the hydrothermal vent environment was crowded with a new generation of daughter cells. With the advent of DNA, primordial life would give rise to the first life

tion. The early development of mRNA-based protocellular life assured genetic continuity by replicating mRNA, mediated by the viral RdRp enzyme, and will be subjected to evolutionary pressure to become more efficient. The protein/RNA protocell, mediated by the RdRp enzyme, is a more realistic start to life than the pure RNA world; it was a likelier precursor to DNA-based life. It was the first attempt to propagate life. With the advent of DNA with high replication fidelity and a permanent depository of genetic information, DNA-based modern cellular life displaced mRNA-based primordial life by competitive exclusion. mRNA-based life became a relict and was lost forever from biogenesis.

We have previously discussed (see Chap. 12) how aminoacyl-RNA (aa-tRNA) created the first gene as a molecular 'transformer,' a term borrowed from the machine language. The sequential nature of RNA/DNA, like that of machine language-processing tasks, is extremely appealing and shows great promise in application in viral and nonviral sequences. Viruses were the second transformers in the origin of life, transforming one nucleic sequence of the protocell to another, thus creating primordial life, and then transforming RNA into DNA sequences. The transformer has been adopted in many natural language-processing tasks but hardly in biological systems, such as RNA/DNA sequences and the origin of life.

15.9 Retroviruses

A retrovirus is a highly derived enveloped particle in which the capsid core contains two identical single-stranded RNA molecules; each carries its genetic and structural blueprint. The virion is 80–100 nm in diameter, and its lipid envelope incorporates and displays viral phospholipids. The hallmark of a retrovirus is its unique replicative strategy, in the sense that it can reverse-transcribe its RNA into DNA using its reverse transcriptase (RT) enzyme. This catalyzed transcription is the reverse flow of information of Central Dogma, hence *reverse transcriptase* and *retrovirus*. The new DNA is then integrated into the host cell genome by an integrase enzyme. The host cell treats the viral DNA as part of its genome, transcribing and translating the viral genes, producing the proteins required to assemble new copies of a retrovirus.

The main virion components of a retrovirus include a phospholipid membrane and two identical double-stranded mRNAs encased in a capsid shell. Each mRNA is typically made up of three genes: the group-specific antigen gene (*gag*), the polymerase gene (*pol*), and the envelope gene (*env*). The *pol* gene encodes the three viral enzymes—protease, reverse transcriptase, and integrase—that catalyze the

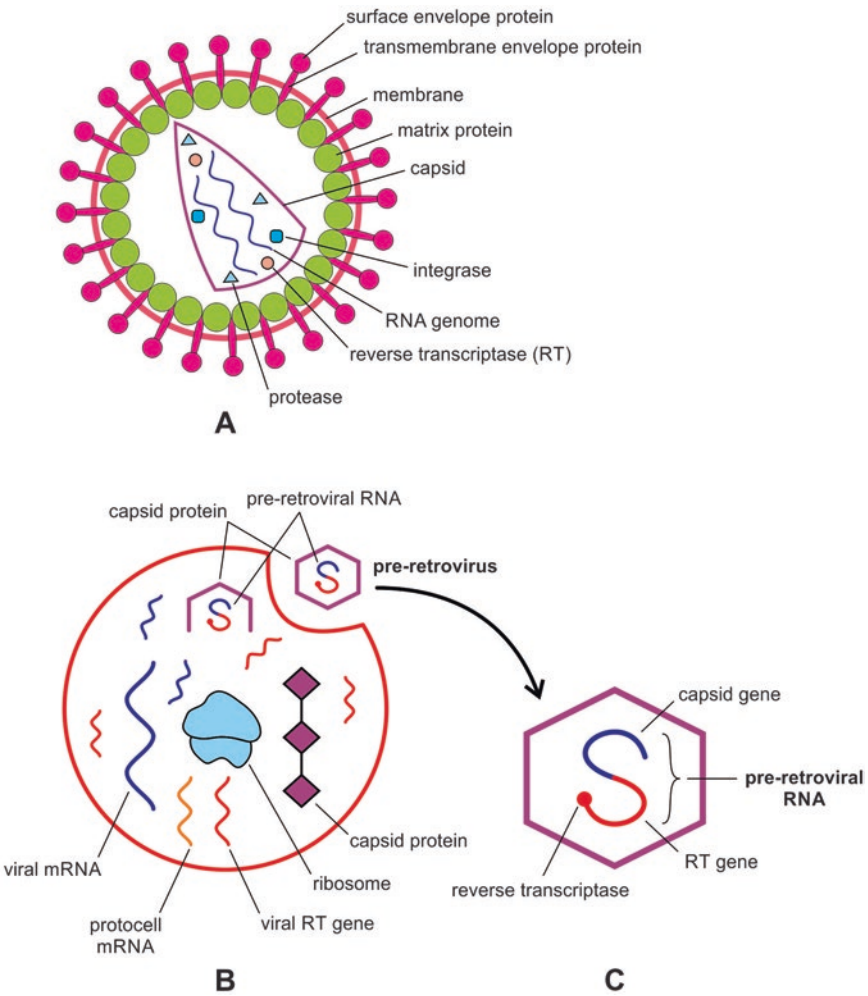


Fig. 15.6 Retroviruses. (**a**) The structure of a modern retrovirus, an enveloped particle in which the capsid core contains two identical single-stranded mRNA molecules. Each mRNA is made up of three genes: integrase, reverse transcriptase, and protease. Once inside the host cell cytoplasm, the virus uses its reverse transcriptase enzyme to produce DNA from its RNA genome. (**b**) The origin of the pre-retrovirus inside an infected protocell, where the viral *RT* gene was linked to the viral *mRNA* gene and encased by a capsid coat. Once two genes were fused into a single hypothetical S-shaped gene for close packing, the pre-retrovirus was released from the protocell via exocytosis. It was a non-enveloped particle. (**c**) The structure of a pre-retrovirus showing how two genes—the capsid gene and the *RT* gene—were fused into a single gene and were encased by a capsid protein

steps of retroviral infection. In retrovirus, multiple protein products are synthesized from a single mRNA species by frameshifting; in between the 5′- and 3′-ends of RNA lies the protein-coding domain, which includes *gag*, *pol*, and *env* encoding regions. These three genes are linked to one another through recoding by frameshifting. The ability to make two or more proteins from the same mRNA is useful, connecting structural (e.g., retroviral *gag*) and catalytic polypeptides (retroviral *pol*). This ribosomal frameshifting mechanism makes the retroviral genome more compact (Fig. 15.6a). Retroviruses have evolved to make use this translational plasticity to regulate their expressions [35].

Once a retrovirus is inside a protocell (a process mediated by protease), it takes over the host's genetic transcription machinery to construct a DNA 'provirus.' Retroviral proteases are critical enzymes in viral propagation and are encoded by a part of the *pol* gene. This process, the conversion of retroviral RNA to proviral DNA, is catalyzed by reverse transcriptase. It is necessary for proviral DNA to integrate into host DNA—a step initiated by the integrase enzyme. Retroviruses can be pathogens of many different hosts, including humans. A notable retrovirus is human HIV, responsible for acquired immunodeficiency syndrome (AIDS).

15.10 The Origin of Retroviruses

The known virosphere consists of three principal viral types: RNA viruses, retroid viruses, and DNA viruses. Horizontal gene transfer (HGT) is rampant among viruses within each of these primary types but is generally confined to closely related viruses or viruses (and plasmids) with similar replication mechanisms [36, 37]. There are many examples of mixing and matching in the virus world, but, mysteriously, they have been confined to the same type of nucleic acid. A novel virus genome discovered in a scorching spring environment suggests recombining two separate groups—an ssDNA virus and an RNA virus—a natural chimera not seen before [38]. In this hybrid genome, alongside the RNA-derived gene, the virus contained a gene for DNA replication typical of a DNA virus. Surprisingly, these hybrid viruses are present not just in acidic lakes but are more widespread in several oceanic samples. This finding proves that modern

viruses combine the information in the two ordinarily separate genetic molecules. Moreover, it supports the idea that viruses performed the upgrading from RNA and effectively gave rise to DNA. Diemer and Stedman [38] suggest that the hybrid virus may have formed when an mRNA virus, a retrovirus, and a DNA virus all infected a cell simultaneously. The retrovirus used its reverse transcriptase enzymes to make a DNA copy of an RNA virus gene, combined with the DNA genome to yield this hybrid. In our view, this hybrid virus provides a first glimpse of the ancient viral birth of DNA in the hydrothermal vent environment by the mixing and matching of an mRNA virus, a pre-retrovirus, and a DNA virus.

Retroviruses are extremely widespread among vertebrates, raising the possibility that they might be as old as their vertebrate hosts. Retroviruses infect many animals, from fish to humans, and can occasionally leave genomic fossils within their host genome, known as endogenous retroviruses (ERVs). ERVs consist of the genetic material of extinct or 'fossil' viruses. Our bodies are littered with shards of retroviruses. In all, 8% of our genome comprises broken and disabled retroviruses, which, millions of years ago, were embedded in our ancestors' DNA. Because they no longer seem to serve a purpose or cause harm, these remains have often been referred to as 'junk DNA' [5]. A recent phylogenetic study of ERVs has placed the time of their most recent common ancestor in the Early Paleozoic [38, 39]. The origin of retroviruses in the Devonian presents a critical framework for investigating evolutionary transitions that led to the emergence of retroviruses. Since vertebrates originated in the Cambrian Sea about 520 million years ago in the Cambrian Period, retroviruses might have developed along with their vertebrate hosts. ERVs in vertebrate hosts represent the upper limit of the retroviral origin. Molecular precursors of retrovirus probably began in the prebiotic environment billions of years ago in the RNA–DNA retro world [7]. Retrovirus-like entities are older than the first cells.

Retroviruses bear many similarities to capsidless selfish genetic elements, such as plasmids and various types of transposons, because they have close evolutionary connections and share hallmark genes. These hallmark genes encode vital components of the viral replication apparatus (such as polymerases and helicases). These retroelements—capsidless genetic parasites—are crucial to understanding the origin of viral genes [26]. We speculate that these retroelements self-assembled to produce a viral gene, such as a positive-strand mRNA virus, step by step. Positive-sense RNAs are particularly suitable for reverse genetics because their genomes are typically infectious in protocells and can be immediately translated by the host's translation machinery. In our model, ancestral retroviruses could emerge only when various retroelements and protein enzymes were available inside protocells.

One of the critical enzymes synthesized inside protocells was the reverse transcriptase (RT) enzyme. RT is an RNA-dependent DNA polymerase. Due to their limited genome size that can be packaged in the virus shell, viral polymerases are generally active as a single protein capable of carrying out multiple functions related to viral genome synthesis. Here, we propose a simple evolutionary scenario for the origin of retroviruses. The large class of retroelements is united by a single conserved gene, the *RT* gene, which defines the critical feature of their replication cycle, reverse transcription [26].

The ancestral stage of a retrovirus is called a pre-retrovirus, or a retroid, a non-enveloped mRNA virus with minimum functional design. Our model derived a pre-retrovirus from the mRNA virus inside an infected protocell. Perhaps, the viral *RT* gene evolved de novo inside the protocell [31]. Most likely, the viral mRNA gene and the viral *RT* gene accidentally merged into one single-stranded mRNA, which was enclosed in a capsid shell. In this fused viral gene, one gene was used to synthesize the structural capsid protein and the other for the *RT* gene was used for the synthesis of the reverse transcriptase enzyme (Fig. 15.6b, c). Pre-retroviruses could use the RdRp enzyme to replicate their genomes.

Making structural proteins and enzymes from the same mRNA gene had distinct selective advantages over two separate genes performing similar functions in a double-stranded RNA virus. Most likely, protocells, by that time, had developed several weapons to ward off viral attacks by destroying viral genomes. In response, the pre-retrovirus altered its RNAs in such a way so as to thwart attacks from the protocells. Moreover, linking two genes (the capsid gene and the *RT* gene) into one mRNA strand allowed a novel, compact, space-saving mechanism of the genome. This ability to link two genes into one was achieved by the ribosomal frameshifting mechanism, as discussed earlier [31]. A pre-retrovirus had now become a single-stranded mRNA virus with a capsid coat to function as a formidable parasite during protocell infection (Fig. 15.6b, c). After the emergence of retroviruses, the protein/RNA world transformed itself into the 'retro world' [26].

15.11 The Origin of DNA Viruses from mRNA Viruses

DNA can be considered modified RNA because there are only two chemical differences between RNA and DNA molecules. The first is removing a single oxygen atom from the ribose of RNA to generate deoxyribose of DNA. The second difference is the addition of a methyl (CH_3) group to the nucleotide base uracil (U) to generate thymine (T) (see Fig. 15.8).

Patrick Forterre [12, 22] suggests three stages in the evolution of DNA from RNA through viral infection: the RNA world, the RNA-to-DNA transition, and the DNA world. In the RNA world, RNA viruses emerged from RNA cells. In this model, LUCA is the primitive RNA cell from which the RNA virus emerged. The RNA virus gave rise to three lineages of the DNA viruses in the RNA–DNA transition. These three lineages of the DNA virus evolved in parallel into three domains of life: bacteria, archaea, and eukaryotes.

Forterre [12, 22, 40] proposes that DNA viruses evolved directly from RNA viruses in two steps during the RNA-to-DNA transition. The viral world supplied the critical enzymes for the RNA-to-DNA conversion. In the first step of the RNA-to-DNA transition, the deoxyribose in DNA was generated from the ribose in RNA by enzymes called ribonucleotide reductases, which converted RNA to U-DNA in the genome. The second step in DNA evolution was converting the uracil base to thymine by thymidylate synthases, forming T-DNA (DNA containing thymidine). The emergence of thymidylate synthase activity in some U-DNA virus lineages produced viruses with the modern form of T-DNA. As these new strains of T-DNA viruses infected protocells, the host genomes gradually transformed from U-DNA to T-DNA. Once deoxyribonucleoside triphosphates were available, their assembly into DNA-like chains followed rapidly. The idea that both ribonucleotide reductases and thymidylate synthases were first encoded in viral genomes and were later transferred to protocells is compatible with phylogenetic analyses of these enzyme families. Thus, viruses donated DNA genomes and replicating genes to the host protocells [12, 22]. The hypothesis of a viral origin for DNA could explain why many DNA viruses encode their ribonucleotide reductase and thymidylate synthase [39]. According to this model, the RNA world and LUCA appeared simultaneously with the existence of a few homologous DNA information proteins in the three cellular domains. Cellular DNA and its replication machinery originated in the DNA world via transfers from DNA viruses to RNA cells. Three such independent transfers led to the origin of bacteria, archaea, and eukaryotes.

Although Koonin et al. [3] accept the viral origin of DNA, they criticize the noncellular LUCA concept of Forterre in the RNA world. Instead, they favor the traditional view that LUCA emerged after the first cells from which all three domains descended. Moreover, they emphasize the crucial role of retroid RNA in the origin of DNA. DNA could not have emerged from the RNA world without reverse transcription. The retro elements must have been among the first classes of primordial viruses that evolved in the primordial genetic pool in the protein/RNA world after the advent of translation, when several kinds of enzymes were synthesized. Integrating such elements into host genomes must have coevolved with the increase in the size of the DNA genome. We find the retroviral infection model of the origin of DNA [7, 27] highly attractive and plausible. In our view, retroid RNA, such as pre-retrovirus, might have invented DNA step by step by converting RNA with the RT enzyme's help. We suggest three stages in the viral evolution, combining both Forterre's and Koonin's models that gave rise to DNA from RNA: the protein/RNA world, the retro world, and the DNA world (Fig. 15.7). In our view, the existence of RNA-only cells (protocells) of Forterre [22] seems to face formidable difficulties. A more parsimonious scenario includes the peptide/RNA world [42].

In the protein/RNA world, viral genes might have evolved de novo in the vent environment from preexisting mRNA. These mRNA genes encoded the JRC capsid protein for a protective shell for stability and durability to form the first virus inside protocells (Fig. 15.3a). These ancient mRNA viruses began to infect protocells to increase their populations in the gene pool (Fig. 15.3b).

In the retro world, pre-retroviruses developed the ability to stitch two genes into one. The first gene encoded the JRC capsid protein, but the second gene encoded the reverse transcriptase (RT) enzyme. The RT enzyme could generate complementary DNA from an mRNA template. The pre-retrovirus would play a crucial role in converting RNA to DNA during recurrent protocell infection and subsequently would be incorporated into the host genome as a provirus. The reverse transcriptase enzyme, also called an RNA-directed DNA polymerase that catalyzed the conversion of RNA to DNA, was not present in the host protocells but was delivered by the pre-retroviruses that converted RNA of the host genome directly into DNA.

The RT enzyme performed three sequential biochemical activities: RNA-dependent DNA polymerase activity, ribonuclease H (RNase H) activity, and DNA-dependent DNA polymerase activity. These activities enabled the RT enzyme to convert single-stranded RNA (ssRNA) into single-stranded DNA (ssDNA). Our model's pre-retroviruses gradually modified their mRNA to DNA during recurrent retroviral infection. As soon as a pre-retrovirus invaded a protocell by endocytosis, its capsid coat was dissolved. The capsid gene segment would be translated into a capsid protein, and the *RT* gene segment into the RT enzyme, thus exploiting the host's ribosomes. The RT enzyme would then transcribe its single-stranded RNA template to single-stranded DNA (ssDNA); the RT would make another strand of cDNA from another mRNA; two strands of cDNA are then combined to create a double-stranded DNA (ssDNA). The conversion of ssDNA to double-stranded DNA was mediated by a DNA polymerase (DdDp) (Fig. 16.1). The information contained in a retroviral gene is thus used to generate the corresponding protein via the sequence:

$$\text{mRNA} \rightarrow \text{DNA} \rightarrow \text{mRNA} \rightarrow \text{protein}.$$

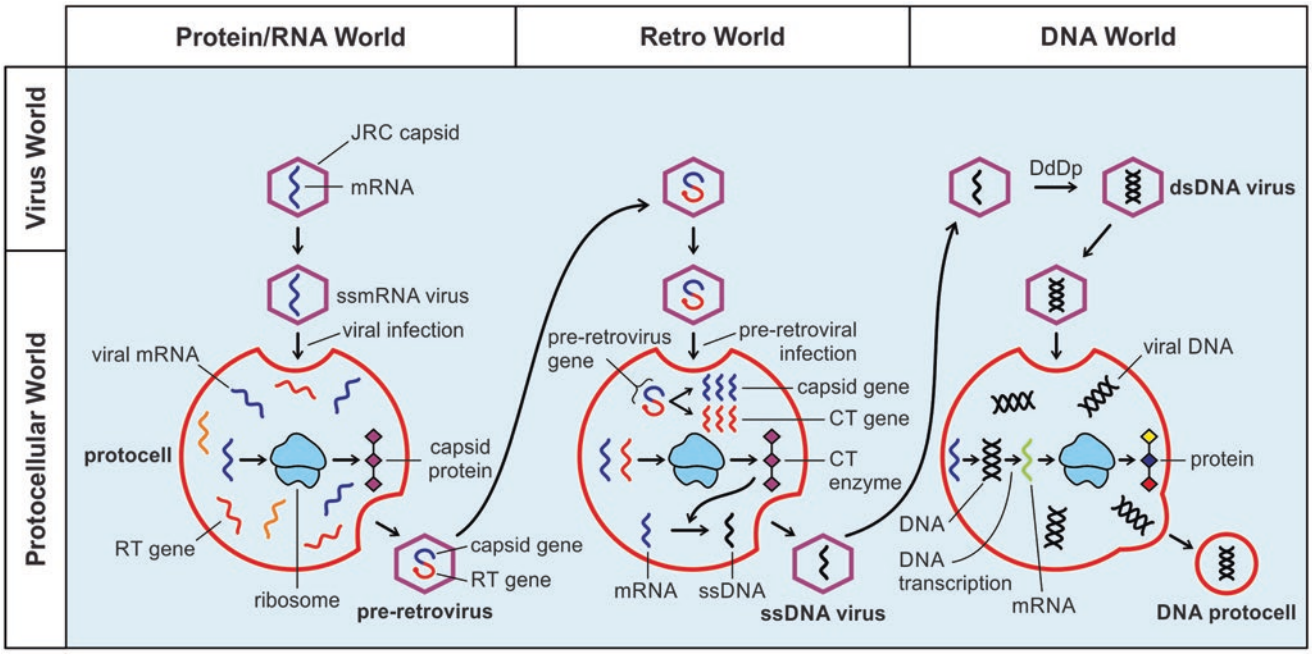


Fig. 15.7 Postulated evolution of the biochemical pathways from the protein/RNA world to the DNA world and the concurrent building of the information system to the contemporary central dogma (DNA makes RNA, and RNA makes protein)

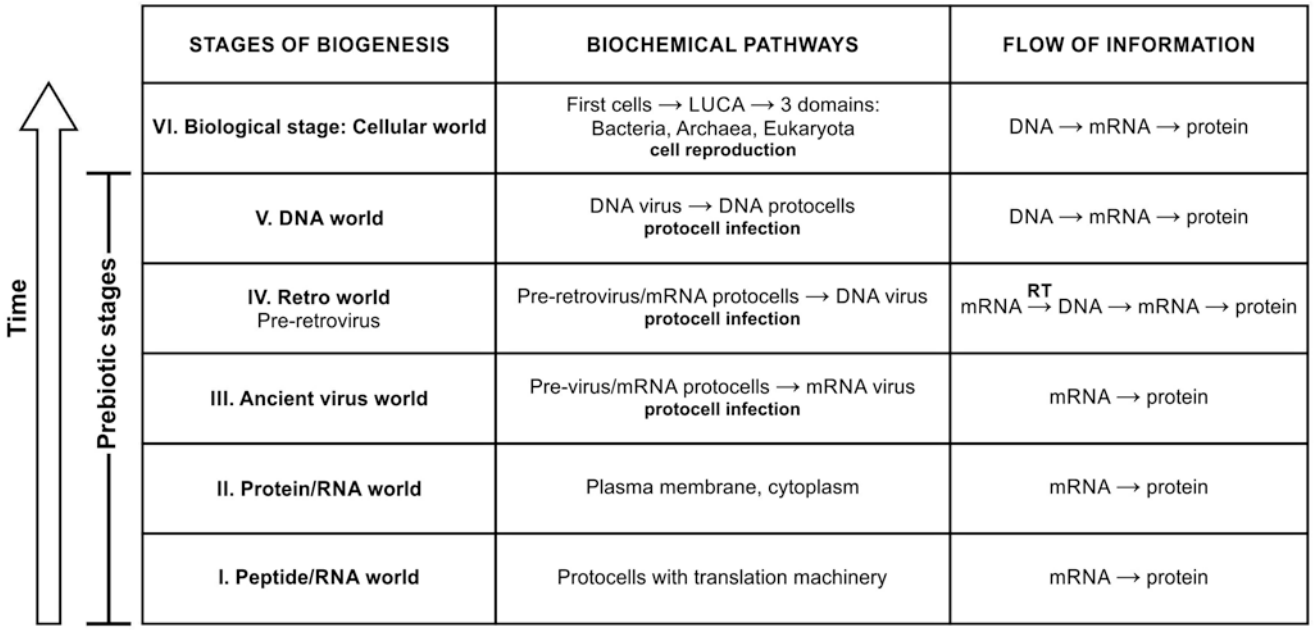
In the DNA world, once the dsDNA virus appeared, the RNA pre-retrovirus's role was gradually replaced by DNA viruses. DNA viruses possessed a complete set of independent DNA replication enzymes: a helicase to unwind the DNA helix, two DNA polymerases (DNA Pols) to replicate the two strands, and a primase to form RNA primers that DNA Pols extended. As dsDNA viruses infected an RNA protocell, its DNA genome was replicated into multiple copies. Each DNA genome replaced the mRNA of the protocell, transcribed to the new generation of mRNA that translated into proteins. This was the beginning of the DNA world when DNA returned mRNA as the protocell's primary genome. The DNA protocell followed a conventional pathway of flow information, as in the central dogma: DNA → mRNA → proteins (Fig. 15.8).

The abundance of genetic parasites and the presence of defense systems in all cellular life-forms suggest that their coevolution might have started in the protocellular stage. DNA viruses might have emerged as a novel survival strategy. When the RNA protocells were confronted with an invading pre-retrovirus, they might have protected themselves by several defense mechanisms and developed an ancestral immune system. The immunity of viral infection allowed the RNA protocells to proliferate. In response, pre-retroviruses might have developed DNA to ward off attacks from the hosts. For pre-retroviruses, DNA might have offered a compelling, immediate benefit: pre-retroviruses might have direct fitness benefits for substituting DNA, replacing its RNA genome [8].

It thus appears that the transition from RNA to DNA genomes occurred in the viral world and that protocellular DNA and its replication machinery originated via transfers from DNA viruses to RNA protocells [11, 40]. The DNA virus living in a carrier state in an RNA protocell probably lost the genes for capsid proteins and became established as DNA plasmids. These plasmids were later transferred to RNA protocells and incrementally transformed into DNA by recurrent infections [22]. With the DNA genome and its transcribed mRNA and ribosome, the RNA protocell modified to the DNA protocell started synthesizing proteins. The coded genetic information began to flow from DNA to RNA to proteins, thus beginning the classical central dogma of molecular biology.

The reverse transcriptase (RT) enzyme that transferred the genetic information from RNA to DNA has disappeared mainly from present-day living cells, where it solely occurs in retroviruses, plasmids, and transposable elements. Similarly, RNA replicases, which presumably played a vital role in replicating RNA genes, have disappeared, except in viruses. This is important concerning the ancient parts of viruses in the evolution of DNA from RNA. Thus, the symbiotic relationship between viruses and protocells is not entirely parasitic but mutualistic. It takes two: how mutualisms evolve in a world of selfish viruses.

There must have been a long transition period during which DNA in a protocell, helped by natural selection, progressively took over the replicative and storage functions of RNA. The DNA genomes are usually far more significant



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

Fig. 15.8 Coevolution of viruses and protocells. Primitive viruses appeared in the protein/RNA world. Since then, viruses have become important vectors for donating critical enzymes and modifying genomes of the protocells during recurrent infection in the origin of life. Later, some of those viruses evolved DNA to defend their genomes from attacks, and DNA-based viruses were incorporated into hosts. In between are the fundamental catalytic processes that allow the stepwise generation of viral deoxyribonucleotides from ribonucleotides by RNA polymerase (RdRp), reverse transcriptase (RT), and DNA polymerase (DdRp) enzymes

than the RNAs, allowing more storage capacity for biological information. Once deoxyribonucleotide triphosphates (dNTPs) were available, their assembly into DNA-like chains would likely have followed rapidly. DNA has intrinsically higher replication fidelity, allowing genomes to increase in size and complexity [25, 41, 42].

15.12 Conclusions

A virus is a hybrid molecule of nucleic acid and protein capsid. It becomes alive and processes digital information during protocell infection by horizontal gene transfer. Gene exchanges between viruses and their hosts are considered the primary drivers of evolution. The advent of proteins heralded the ancient virus world when viral genes began concentrating in hydrothermal vent environments and were randomly coated by proteins for durability and protection. The naked viral world and the compartmentalized protocellular world evolved in the protein/RNA world. Viral genes have two components, one for making capsid proteins and the other for replicating nucleic acids. These primitive viruses began infecting protocells and exploiting their ribosomes to create viral proteins. This was the beginning of mRNA viruses. Once the capsid protein began to coat the viral gene, the first mRNA virus appeared. These mRNA viruses began to infect protein/RNA protocells by horizontal gene transfer (HGT) and horizontal enzyme transfer (HET), such as RNA-dependent RNA polymerases (RdRps). These primitive RNA viruses began to infect protocells by HGT and HET and exploited their ribosomes for reproduction. The protein/RNA protocell, in turn, became the first primordial life capable of protocellular reproduction. The next stage in viral evolution was the emergence of a primitive retrovirus with a new kind of replicative strategy that could turn its RNA into DNA using its own reverse transcriptase (RT) enzyme that facilitated the transition from RNA to DNA genomes. With persistent infection, DNA viruses slowly transfer to protocells, with not only their core replication enzymes by HET but also their DNAs by HGT. Thus, the symbiotic relationship between viruses and protocells is not entirely parasitic but mutualistic. Viruses were truly ‘transformers’ in the origin of life.

Thus, began the ‘DNA world’ when DNA replaced RNA as the major genome of the protocells. As a planetary force for life, viruses account for many of evolution’s most exceptional selective pressure and innovations. Viruses are ubiquitous companions of cellular life-forms and play a crucial role in the origin of life. The existence of several genes that are central to viral replication and structure are shared by many groups of RNA and DNA viruses but are missing in the cellular world. Virus hallmark genes suggest the concept of the ancient virus world, a flow of virus-specific genes that has continued uninterrupted from the precellular stage of life’s evolution to this day.

References

1. La Scola B, Audic S, Robert C, Jungang L, Lamballerie XD, Drancourt M, Birtles R, Claverie JM, Raoult D (2003) A giant virus in Amoebae. Science 299:2033
2. Forterre P (2010) Defining a life: the virus viewpoint. Orig Life Evol Biosph 40:151–160
3. Koonin EV, Senkevich TG, Dojla VV (2006) The ancient virus world and evolution of cells. Biol Direct 1:29. https://doi.org/10.1186/1745-6150-1-29
4. Suttle CA (2007) Marine viruses—major players in the global ecosystem. Nat Rev Microbiol 5:801–812
5. Koonin EV (2016) Viruses and mobiles elements as drivers of evolutionary transitions. Philos Trans R Soc B 371. https://doi.org/10.1098/rstb.2015.0442
6. Villarreall LP (2005) Viruses and evolution of life. American Society of Microbiology Press, Washington, DC
7. Wagner EK, Hewlett MJ (2004) Basic virology. Blackwell Publishing, Malden
8. Villarreall LP (2004) Are viruses alive? Sci Am 291(1):100–105
9. Jaimes JA, Andre NM, Chappie JS, Millet JK, Whittaker GR (2020) Phylogenetic analysis and structural modeling of SARS-CoV2 spike protein reveals and evolutionary distinct and proteolytically sensitive activation loop. J Mol Biol 432:3309–3325
10. D'Herelle F (1922) The bacteriophage: its role in immunity. Williams and Wilkins, Baltimore
11. Haldane JBS (1929) The origin of life. Ration Ann 148–169:3–10
12. Forterre P (2005) The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. Biochemie 87:793–803
13. Margulis L (2005) Symbiosis in cell division. Freeman, San Francisco
14. Takemura M (2001) Poxyviruses and the origin of the eukaryotic nucleus. Mol Evol Biol 52:419–425
15. Chatterjee S (2009) The river of life: a genetic perspective on macroevolution. Forum Public Policy J Oxf Round Table 5(5):1–43
16. Doolittle WF (2000) Uprooting the tree of life. Sci Am 282(2):90–95
17. Daubin V, Ochman H (2004) Start-up entities in the origin of new genes. Curr Opin Genet Dev 14:616–619
18. Feschotte C (2010) Virology: bornavirus enters the genome. Nature 463:39–40
19. Chuong EB (2013) Retroviruses facilitate the rapid evolution of the mammalian placenta. Bioessays 35(10). https://doi.org/10.1002/bies.201300059
20. McLaughlin RN, Malik HS (2017) Genetic conflicts: the usual suspects and beyond. J Exp Biol 20:6–17
21. Holmes EC (2011) What does virus evolution tell us about virus origin? J Virol 85:5247–5251
22. Forterre P (2006) The origin of viruses and their possible roles in major evolutionary transitions. Virus Res 117:5–16
23. Koonin EV (2009) On the origin of cells and viruses: primordial virus world scenario. Ann N Y Acad Sci 1178:47–64
24. Chatterjee S (2016) A symbiotic view of the origin of life at hydrothermal impact crater lakes. Phys Chem Chem Phys 18:20033–20046
25. De Duve C (2005) Singularities: landmarks on the Pathways of Life. Cambridge University Press, New York
26. Koonin EV, Dojla VV (2013) A virocentric perspective on the evolution of life. Curr Opin Virol 3(5):546–557
27. Koonin EV, Dojla VV (2014) Virus world as an evolutionary network of viruses and capsidless selfish elements. Microbiol Mol Biol Rev 78:278–303
28. Krupovic M, Koonin EV (2011) Multiple origins of viral capsid proteins from cellular ancestors. Proc Natl Acad Sci. https://doi.org/10.1073/pnas.1621061114
29. Filee J, Forterre P (2005) Viral proteins functioning in organelles: a cryptic origin? Trends Microbiol 13:510–513
30. Durzynska J, Gozdzicka-Jozefiak A (2015) Viruses and cells intertwined since the dawn of evolution. Virol J 12:169. https://doi.org/10.1186/s12985-0400-7
31. Sabbath N, Wagner A, Karlin D (2012) Evolution of vital proteins originated de novo by overprinting. Mol Biol Evol 29:3767–3780
32. Joyce GF, Szostak JW (2018) Protocells and RNA self-replication. Cold Spring Harb Perspect Biol 2018(10):a034801
33. Kockler ZW, Gordenin DA (2021) From RNA world to SARS-Co V-2: the edited story of RNA viral evolution. Cell 10:1557. https://doi.org/10.3390/cells10061557
34. Drake JW, Holland JJ (1999) Mutation rates among RNA viruses. Proc Natl Acad Sci U S A 96:13910–13913
35. Ahlquist P, Noueiry AO, Lee WM, Kushner DB, Dye BT (2003) Host factors in positive-strand RNA virus genome replication. J Virol 77:8181–8186
36. Brierly I (1995) Ribosomal frameshifting on viral RNAs. J Gen Virol 76:1885–1892
37. Krupovic M, Prangishvilli D, Hendrix RW, Bamford DH (2011) Genomic of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. Microbiol Mol Biol Rev 75:610–635
38. Diemer GS, Stedman KM (2017) A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. Biol Direct 7:13. https://doi.org/10.1186/1745-6150-7-13
39. Aiewsakun P, Katzourakis A (2017) Marine origin of retroviruses in the Early Paleozoic Era. Nature Commun 8:13954. https://doi.org/10.1038/ncomms133954
40. Forterre P (2006) Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. Proc Natl Acad Sci 103:3669–3674
41. Powner MW, Zheng SL, Szostak JW (2012) Multicomponent assembly of proposed DNA precursors in water. J Am Chem Soc 134:13889–13895
42. Chatterjee S, Yadav S (2019) The origin of prebiotic information system in the peptide/RNA world: A simulation model of the evolution of translation and the genetic code. Life 9:25. https://doi.org/10.3390/life9010025

16 DNA Takes Over

DNA is like a computer program but far, far more advanced than any software ever created.
—Bill Gates, 1995

16.1 The Birth of DNA

Retroviruses contributed to the origin of DNA by horizontal gene transfer (HGT) and horizontal enzyme transfer (HET) through recurrent viral infections of protein/RNA protocells. DNA acquired two important functions, namely, transcription and replication, from DNA viruses, using their viral enzymes. As the genetic memory, diversity, and sophistication increased in protocells, DNA replaced RNA owing to the former's stability and increased storage capacity for genetic information.

The upgrading from RNA to the DNA world was a significant event in life history [1]. The passage from an RNA genome to a DNA genome was relatively simple with the aid of retroviruses. These two nucleic acids are so similar that the change of RNA to DNA requires two components of RNA to be replaced by two close relatives: ribose by deoxyribose and uracil (U) by thymine (T). A set of enzymes assisted both processes. RNA evolved into DNA by reverse transcription, mediated by retroviruses. DNA's stable code can encrypt a considerable amount of digital information. We have already discussed the DNA protocell's origin from the RNA protocell by pre-retroviruses (Fig. 15.6). Except for RNA retroviruses, believed to be relics of the retro world (Fig. 15.7), RNA replicase and reverse transcriptase have largely disappeared from the biosphere. Nature left the riddle of the birth of DNA hidden in the retro world.

Although the transformation of DNA from RNA via retroviruses is widely accepted [2, 3], there are some dissenters. The latest twist in the origin of the DNA debate is that RNA and DNA might have emerged simultaneously in the prebiotic world from building blocks without retroviral assistance. Powner et al. [4] proposed a novel pathway for the prebiotic synthesis of several molecular components for making DNA molecules from a mixture of the chemicals available in the sulfur-rich prebiotic environment, where both RNAs and DNAs emerged simultaneously, not one after another. They argued that transforming RNA nucleotides into DNA nucleotides required special energy-rich enzymes. In contrast, if DNA and RNA molecules were present simultaneously, then this problem of radical switching from RNA to DNA could be solved. The abiotic route for the synthesis of deoxyribonucleic acid provides a new perspective for the origin of DNA.

Other authors have expressed a similar view. For example, Xu et al. [5] suggested the de novo origin of DNA from the prebiotic building blocks of life. They argued that thiouridine, presumably present in the prebiotic environment, could have joined DNA nucleotides to form chain-like DNA. Prebiotic phosphorylation of the 2-thiouridine molecule produced DNA nucleotides via photoreduction. Both RNA and DNA may have arisen together during prebiotic synthesis.

The synchronous origin of DNA and RNA appears to be a less parsimonious explanation than the hypothesis that RNA preceded DNA and proteins in evolution. DNA molecules must have evolved later than RNA molecules because, unlike RNA, DNA could not emerge without the availability of several enzymes. RNA degrades and mutates easily [6]. The backbone of single-stranded RNA is much less stable than the equivalent structure of double-stranded DNA. The enhanced stability and more extended molecular sequence provide DNA with higher fidelity and increased memory in its information storage system. RNA replication is intrinsically error-prone compared with DNA replication. The selection of DNA over RNA was based on its expanded capacity to store information and dramatically improved error rate during replication [7].

The recent detection of ribose sugar in the Murchison meteorite and other primitive carbonaceous chondrites (NWA 801 and NWA 7020) by an international team of scientists has suggested that RNA evolved before DNA, not concurrently [8]. Ribose is a crucial component of RNA, which could have stored information and catalyzed reactions

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_16

during prebiotic synthesis. The research provides the first direct evidence of ribose in space and the delivery of the sugar to Earth during the impacts of carbonaceous meteorites about four billion years ago. Remarkably, a molecule as fragile as ribose could be detected in such ancient meteorites. In contrast, the sugar in DNA (deoxyribose) was not recognized in any of the meteorites analyzed in this study. This finding is important since there could have been a delivery bias of extraterrestrial ribose to early Earth, consistent with the hypothesis that RNA evolved first. According to the team, the next logical step would be to investigate the chirality of sugars in more carbonaceous chondrites to check whether the cosmic sugar is right-handed. Perhaps, homochirality of extraterrestrial ribose evolved on the mineral surface in the hydrothermal vent environment, not in space [9]. The extraterrestrial ribose might have contributed to the formation of RNA on prebiotic Earth, possibly leading to the origin of DNA by retroviral infection [3, 10].

16.2 The Evolutionary Advantage of DNA Over RNA

As the genetic diversity increased and the translation perfected, DNA protocells must have faced dilemmas: how to expand the information storage capacity of mRNA and how to separate translation from replication. Unlike RNA, DNA is a poor catalyst. RNA can act as both a template and a catalyst, whereas DNA can function only as a template. The division of labor between the RNA catalyst and the DNA template could solve this problem. Information storage and replication is the prerogative of DNA, whereas the utilization of this information for protein synthesis and other functions remains the province of RNA. When replication of information is entirely dissociated from its expression, DNA took on the role of a protocell. Because genes no longer had to serve messengers, they could be joined together in strings of increasing length to maximize the storage capacity of information permanently [1, 11]. Moreover, mRNA could be transcribed from DNA, when necessary and expressed during protein synthesis. RNA is more versatile, acting in both the storage and replication of genetic information.

However, during the transition from RNA to DNA, the DNA lost its catalytic power but transformed into a stable, long-term storage molecule for genetic data. The functional separation between replication and catalysis must have resulted in a tremendous improvement in DNA molecules and the functionality of protocells. Strong selective forces favored it. The division of labor between a template and a catalyst is a fundamental attribute of all living systems. This underlying life property did not appear in the protein/RNA world but later did so in the DNA world. Another advantage of DNA is that it allows a selective expression of individual genes by transcription, thus keeping the DNA molecule intact. In contrast, after translation, each mRNA molecule is destroyed.

With the advent of DNA, all the genes could now be kept as stable, double-stranded threads, which are held together by complementary base pairing and are twisted in a double helix. The extremely durable structure does not allow DNA to mutate rapidly, unlike RNA, thus converting to an efficient information storage structure. Intact stretches of DNA have been recovered from fossil bones that are at least 700,000 years old [12]. The DNA acts as a permanent record: a blueprint containing the information needed to build proteins, run cellular functions, and reproduce life. In transcription, a gene, a segment of DNA, is transcribed to make a short-lived mRNA molecule or a gene for translation, but DNA preserves the information storage system. Storage in mRNA without the possibility of retrieval would have been useless, and, hence, there is a need for transcription. The stored information could thereby be recovered from mRNA for translation. Transcription becomes entirely dissociated from translation during the division of labor between DNA and RNA.

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

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

The RNA is much more susceptible to chemical transformation. The DNA backbone is less prone to hydrolysis because it lacks the nucleophilic 2-hydroxyl group, so it represents a more chemically stable material and is less prone to mutation. As a result, DNA replicates with higher intrinsic fidelity than does RNA, thus allowing more information storage. An RNA polymerase, the enzyme that generates RNA from a DNA template, exhibits no proofreading activity, and

whereas DNA has many repair mechanisms, RNA has none. DNA soon superseded RNA as the carrier of genes and became the dominant bearer of genetic information. Being much longer than RNA, DNA can store information on thousands of genes. Once DNA was established, it allowed genes to become longer and more complex. The fidelity of DNA replication is orders of magnitude higher than that of RNA replication. Unlike a single-stranded RNA molecule, in the DNA molecule, 2 antiparallel strands that are complementary in their nucleotide sequences are paired in a right-handed double helix, with about 10 nucleotide pairs per helical turn (Fig. 14.5).

Once DNA molecules appeared on the scene in the DNA world, they took on the role of the primary genetic molecules, superseding the RNA molecules, which became intermediaries between DNA and proteins. These new DNA-containing protocells rapidly diversified into large populations that easily outcompeted the RNA-based protocells [13, 14].

16.3 DNA Structure

Structurally, the difference between DNA and RNA lies in their strandedness. Apart from some viral forms, all RNAs are single-stranded; an RNA's base pairs mostly with itself to form loops closed by short, double helical complementary segments to form a hairpin structure or a transfer RNA (tRNA). In contrast, all DNAs are made of two complementary antiparallel strands, except for a few small viruses. Double-stranded DNA is more rigid than single-stranded RNA.

The double helix structure of DNA, proposed by Watson and Crick [15], an iconic image, is based on two paired DNA strands complementary in their nucleotide sequence. The 'immortal coil' of two intertwined strands creates a molecule with the shape of a twisted ladder. The sides of the ladder are made of the sugar–phosphate backbones of the two strands. The four nitrogen bases are adenine (A), thymine (T), guanine (G), and cytosine (C). These bases project off the sugar–phosphate backbone and join together by hydrogen bonds, forming the 'rungs' of the ladder (Fig. 15.8). Each strand of DNA is a long sequence of four bases. The order of these bases is what determines DNA's instructions or genetic code. Both strands of double-stranded DNA store the same biological information. The two chains are held together by hydrogen bonds. A pairs with T by forming two hydrogen bonds, whereas G pairs with C by forming three hydrogen bonds. Because the A–T and G–C pairs are equal in length and fit identically into the double helix (like rungs on a ladder), the helix's diameter is uniform, i.e., 2.0 nm (Fig. 16.1). These two strands are complementary, but they go in the opposite or antiparallel direction and twist helically. the sugar-phosphate backbone is negatively charged and hydrophilic, which allows DNA backbone to form bonds with water. The double helix of DNA's two strands is held together by bases pairing in the antiparallel orientation. The spiral is 'right-handed'—twisting upward as if driven by a right-handed screw, a chiral feature. Some scientists believe that cosmic rays may have given right-handed genetic helixes an evolutionary edge during prebiotic synthesis. The DNA sequence directly encodes each protein's structure, which determines its activity. All these features increase DNA's stability and its effectiveness as a reliable information-bearing molecule [14].

Solving DNA structure immediately revealed how the two fundamental processes of inheritance and mutation worked at a molecular level [14, 15]. A DNA sequence could be faithfully copied and passed on because the base present at each position on one strand determined its complement on the other strand. Mutations result from errors in copying processes, in which the wrong base or extra bases get inserted, or a base may be deleted, thus generating a change in the DNA sequence. The error rate of information-processing transactions of DNA is about one error in a billion base-pair replications.

Two critical functions evolved with the advent of DNA: transcription and replication. Transcription is the first step in gene expression. It involves copying a gene of the DNA sequence to make an mRNA molecule for protein synthesis. Replication, on the other hand, is the process by which a double-stranded DNA molecule is copied to produce two identical DNA molecules. As discussed earlier, viruses provided the critical enzymes needed for transcription and replication during recurrent infection of protocells [3].

16.4 DNA Transcription

DNA is generally regarded as the memory mapping of a protein structure. Yet, it is mRNA that is transcripted from DNA as the input and a protein is the output. The input of an event disappears automatically with the output of the event. This is why after translation, mRNA is recycled but DNA remains intact. The storage of information in DNA without the possibility of retrieval would have been useless. RNA can function as both a template-directed polymerase and a template, whereas DNA can function only as a template. As their genetic diversity and sophistication increased, DNA protocells needed a division of labor between the DNA template and its catalyst mRNA. In other words, replication had to be separated from translation [11]. Therefore, a transcription mechanism evolved with the emergence of DNA. The information stored in DNA must be recoverable in mRNA for translation. Transcription is the first step in DNA-based protein synthesis. A particular segment of the DNA template is

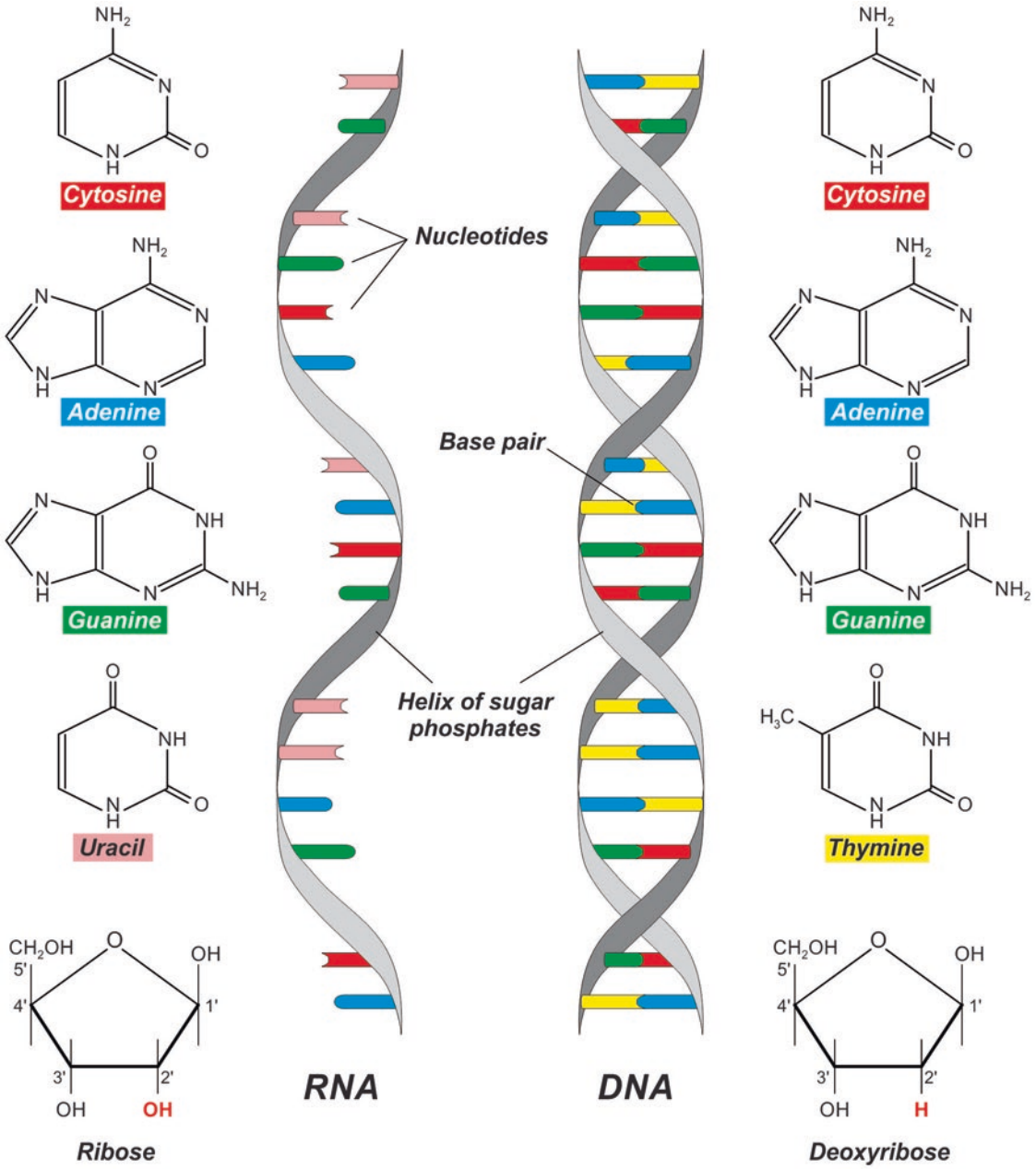


Fig. 16.1 Comparison of RNA and DNA showing the differences in their structures and their nucleobases. RNA is single-stranded, but DNA is double-stranded. RNA contains the sugar ribose, whereas DNA contains the sugar deoxyribose. Ribose has one more –OH group than deoxyribose, with –H attached to the second (2′) carbon in the ring. RNA and DNA base pairing is slightly different since DNA uses the thymine, whereas RNA uses uracil. Uracil differs from thymine, in that it lacks a methyl group

copied into a complementary mRNA by the DNA-dependent RNA polymerase (DdRp) enzyme, thus transferring genetic information from a section of the DNA into mRNA. Transcription proceeds in the 5′- to 3′-direction. That transcript, an mRNA gene, is then used to produce a protein.

Each gene codes for a particular protein. Although genes usually code for proteins, they do sometimes code for RNA. They code all the information necessary to make an organism. Transcription is the synthesis of mRNA from a DNA template. It is the first step in the synthesis of proteins. Transcription begins when a special enzyme, RNA polymerase, binds to a particular sequence of nucleotides on a strand of the DNA. Just as the promoter site signals to the polymerase where to begin, so does a ‘stop’ signal at the far end of the gene signal the polymerase to stop transcription.

The transcription process can be broadly categorized into three main stages: *initiation*, *elongation*, and *termination*. In the initiation stage, RNA polymerase catalyzes transcription.

It slides along the DNA molecule until it recognizes a promoter sequence, indicating the starting point of transcription. Once bound to the promoter sequence, RNA polymerase unwinds a portion of the DNA helix, separating the bases of the two DNA strands.

In the elongation stage, one DNA strand (the template strand) is read in a 3′- to 5′-direction and provides the template for the new mRNA molecule. The other DNA strand is referred to as the coding strand. Incoming ribonucleotides are used to form the mRNA strand. It does this by complementary base pairing. RNA polymerase then catalyzes the formation of phosphodiester bonds between adjacent nucleotides. It travels along the length of the strand of DNA and binds complementary RNA nucleotides to it until a complete strand of mRNA is formed, encoding at least one gene (Fig. 16.2a).

In the termination stage, elongation will continue until the RNA polymerase encounters a stop sequence. Once a gene is transcribed, the newly made mRNA is dissociated from the DNA template. As the mRNA molecule is formed, the DNA helix zips itself back together. This new generation of mRNA is then translated by a ribosome to synthesize a protein chain.

In the next stage of gene expression, the mRNA, using its copy of the DNA code, directs the construction of a polypeptide. This protein synthesis process is called translation because the data stored as nucleotides in DNA and mRNA are translated into an amino acid code to form proteins. In this way, the nucleotide sequence of information is translated into amino acid sequence information. At this time, the central dogma emerged, combining the two-step process, transcription, and translation, by which the information flows from DNA to mRNA to proteins (Fig. 16.2b) [16]. Although

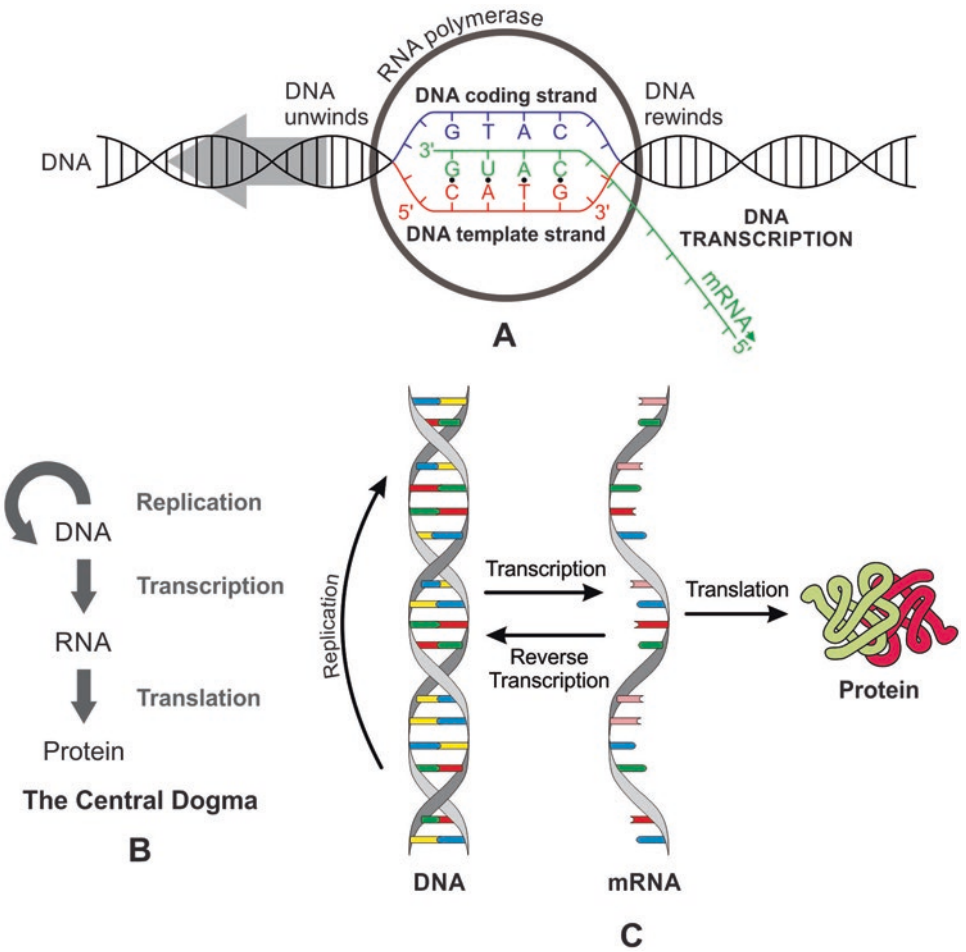


Fig. 16.2 DNA transcription, mRNA translation, and the central dogma. In transcription, a gene is transcribed to a newly assembled piece of mRNA. It is mediated by the RNA polymerase and transcription factors. (**a**) Protein synthesis begins when a region of DNA is split apart, and a molecule of mRNA is built along one template strand by an enzyme called the 'RNA polymerase.' When the mRNA transcript is formed, it peels away from the DNA, allowing the already transcribed DNA to rewind into a double helix. Eventually, the matured mRNA finds its way to a ribosome, where it is translated into a particular protein. (**b**) The genetic information flows in one direction when genes are expressed from DNA to RNA to proteins (DNA → RNA → proteins), called the central dogma. (**c**) However, the retroviruses transcribe mRNA to DNA through the special enzyme reverse transcriptase (RT), resulting in an exception to the central dogma: RNA → DNA → RNA → proteins

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

Reverse transcription is the transfer of information from RNA to DNA (the reverse of standard transcription), as in retroviruses, using a special enzyme called reverse transcriptase, thus providing alternative pathways for the central dogma (Fig. 16.2c). However, this reverse transcription gave rise to DNA from RNA during prebiotic synthesis and was a critical stage in abiogenesis. Reverse transcription is an essential step in human immunodeficiency virus 1 (HIV-1) infection. The discovery of reverse transcriptase, which won a Nobel Prize in 1970 for its discoverers, Howard Temin and David Baltimore, provided an invaluable tool for biotechnology.

16.5 DNA Replication

DNA replication is a polymerization process that requires a template and a primer, and the product of the reaction is a new strand of DNA that is complementary to the template strand. For successful reproduction, a cell must be able to copy and transmit all of its genetic information stored in DNA to each of its daughter cells. The structure of DNA allows cells to do this. DNA replication is central to the reproduction of all cellular life forms and many viruses [14]. Every time a cell divides, DNA polymerases are required to help duplicate the cell's DNA so that a copy of the original DNA molecules can be passed to each daughter cell. In this way, genetic information is passed from generation to generation. Cellular DNA replication systems are broadly classified into two types, one in bacteria and the other in archaea/eukaryotes. In contrast, double-stranded DNA viruses have a much broader diversity of DNA replication systems. Both Koonin [10] and Forterre [17] suggest that protocells acquired the DNA replication core enzymes such as helicase, primase, and DNA polymerases (Pols) from DNA viruses during their recurrent infection of protocells by horizontal enzyme transfer (HET).

The double helix structure of DNA had striking implications for the process of DNA replication. There are some similarities between transcription and replication. In both cases, the DNA double helix is untwisted when the hydrogen bonds between the bases are broken. However, there are significant differences between these two processes. Transcription copies the DNA into the mRNA, whereas replication makes another copy of the DNA. In transcription, strand separation is mediated by the RNA polymerase, whereas in replication, the DNA polymerase takes up this role. Both processes involve the generation of a new molecule of nucleic acid, either mRNA or DNA. However, each process's function is extremely different, with one involved in gene expression and the other in cell division. Transcription was the precursor to replication. DNA replication, on the other hand, must have evolved as a prelude to cell division.

The DNA replication process is well documented in the literature [14]. Its basic idea will be discussed briefly to highlight its role in the first cell division and the origin of life. The antiparallel structure of DNA is essential in DNA replication because it replicates the leading strand in one way and the lagging strand the other way. DNA replication is performed by a replisome, a complex molecular machine composed of numerous enzymes. It is an elaborate process that requires the coordinated effort of a team of enzymes for unwinding, separation, replication, and rewinding the helix. The replisome comprises two replicative polymerase complexes: one synthesizes the leading strand, whereas the other synthesizes the lagging strand.

The following enzymes are involved in DNA replication:

Helicase breaks or melts the hydrogen bonds that hold the coiled helix's two halves together to unwind the DNA duplex.

Topoisomerase breaks and reseals the DNA backbone to allow the backbone to release tension caused by twisting.

Primase, also called RNA polymerase, puts down tiny pieces of RNA called *primers* that DNA Pols extend. The primers serve as starting points for copying the DNA.

DNA polymerase makes new strands of DNA replication: Two different DNA polymerases (Pols) are involved in DNA replication:

- *DNA polymerase III* copies the DNA by reading the code on the existing strands complementary to the originals. The enzyme follows the base pairing rules.
- *DNA polymerase I* removes RNA nucleotides from the RNA primers and replaces them with DNA nucleotides so that no RNA is remaining in the DNA when the job is done.

DNA ligase seals up breaks in the DNA backbone by forming covalent bonds between nucleotides.

The enzymes involved in DNA replication stay together in a large enzyme complex. All enzymes work simultaneously, each doing their part in the sequence of events that needs to occur to copy the DNA. Replication occurs in three significant steps: the opening of the double helix and separation of the DNA strands, the priming of the template strand, and the assembly of the new DNA segment.

During separation, the two strands of the DNA's double helix uncoil at a specific location called the origin. The initiation of the DNA replication occurs when an initiator enzyme called a helicase unwinds a short stretch of the DNA double helix using ATP hydrolysis's energy. It breaks apart the hydrogen bonds between the bases of the DNA strands. It

opens up like a zipper in one direction at the Y junction, or replication fork, to form two strands: a continuous leading strand and a small, discontinuous lagging strand. In DNA, these two strands are antiparallel; the 3′-end of the leading strand is paired with the 5′-end of the lagging strand. Each strand serves as a template for a new strand (Fig. 16.3c).

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

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

The synthesis of the lagging strand is more complicated because the DNA strand here is antiparallel. The DNA polymerase must work in the opposite of the replication fork in the 5′ → 3′-chemical direction. The synthesis of the lagging strand starts when a primase synthesizes a short stretch of RNA that acts as a primer. Synthesis requires repeated priming and extension of the lagging strand discontinuously as a series of Okazaki fragments (Fig. 16.3). The DNA polymerase then adds bases to the 3′-end of the lagging strand. Here, the synthesis is discontinuous, where the DNA polymerase can synthesize relatively small, discontinuous stretches in the 3′ → 5′-direction. These short stretches of new DNA are called Okazaki fragments, which are then linked by the ligase enzyme to form a continuous whole. Once all the bases are matched up, an enzyme called ribonuclease H strips away the RNA primers. Finally, an enzyme called the DNA ligase seals up the sequences of the DNA into two continuous double strands.

The result of DNA replication is two DNA molecules consisting of one new and one old chain of nucleotides. Therefore, DNA replication is described as semiconservative; half of the chain is part of the original molecule, half is brand new. Following replication, the new DNA is automatically coiled tightly around an axis to form a double helix for stability and greater compactness. In bacteria, circular DNA is a higher-order helix-upon-a-helix, known as a superhelix (Fig. 16.3).

The DNA replication apparatus has evolved over billions of years through ‘trial and error’ since its inception in the protocell in the DNA world. The elaborate process by which DNA is replicated today could not have been the original version of replication, so a simpler, more primitive mechanism remains to be discovered. One possibility is that in the DNA genome of the protocell was an extremely small number, perhaps just a few hundreds, of nucleotides, where replication was less cumbersome, requiring few unwinding and rewinding enzymes and DNA polymerases for replication of each strand (Fig. 16.3). It required two replisomes for bidirectional replication. These short nucleobases of two DNA strands were uncoiled and completely separated into two strands; each strand can serve as a template for making a new complementary strand. The net result was the formation of two new double-stranded DNA sequences that were exact copies of the original double-stranded DNA. In this way, the double helical DNA could be copied precisely.

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

In a cell, all RNAs are created by transcription, a process that has certain similarities to the process of DNA replication. Transcription, however, differs from DNA replication in several crucial ways. Unlike a newly formed DNA strand, the RNA strand does not remain hydrogen-bonded to the DNA template strand. Instead, the RNA chain is being displaced just behind the region where the ribonucleotides are added, and the DNA helix reforms. Thus, the product by transcription is released from the DNA template as a single strand. Moreover, because they are copied from only a limited region of DNA, RNA molecules are much shorter than DNA molecules. The two enzymes used for making copies are also quite different. The enzymes that perform transcription are RNA polymerases that catalyze the transcription by forming the phosphodiester bonds that link the nucleotides together to form a linear chain.

16.6 The Flow of Genetic Information

With the advent of DNA, the basic flow of genetic information in biological systems, often depicted in a scheme known as the ‘central dogma,’ was established [16]. It states that DNA makes RNA and RNA makes proteins (Fig. 16.2b). The general statement of the central dogma suggests that the

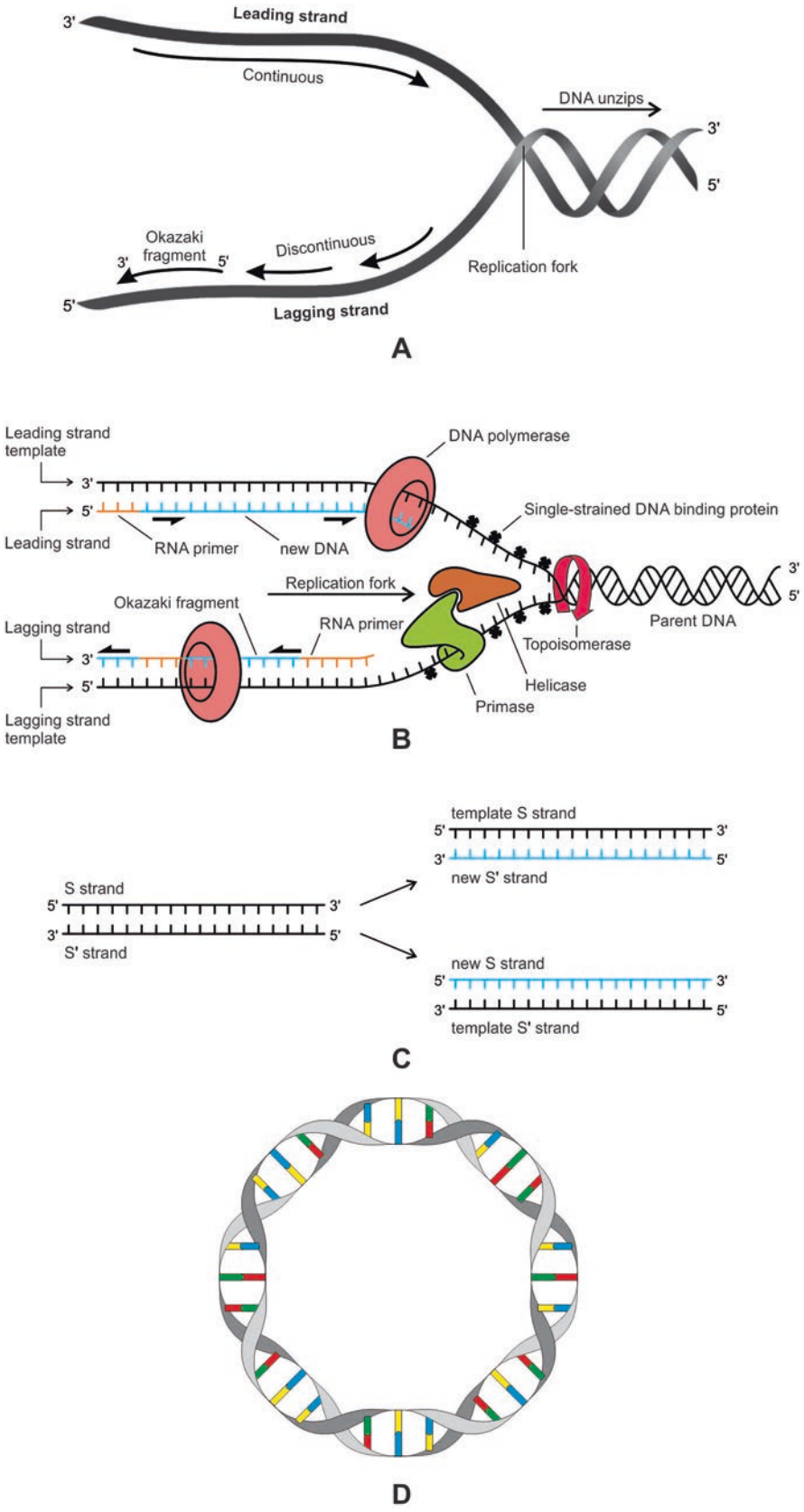


Fig. 16.3 Replication of DNA by DNA polymerase (Pols). (**a**) To begin DNA replication, the DNA helicase enzyme causes the two parent DNA strands to unwind and separate from one another to form a Y-shaped replication fork. Both new strands are synthesized in the 5′- to 3′-direction. The leading strand grows continuously forward, but the lagging strand grows in short discontinuous stretches called Okazaki fragments. (**b**) Many core enzymes—helicase, primase, and DNA polymerases—collaborate at the replication fork and are involved in DNA replication. (**c**) Primordial DNA replication in a protocell. Here, the short nucleobases of DNA are uncoiled and completely separated into two strands as S and S′; S can serve as a template for making a new strand S′. In contrast, strand S′ can serve as a template for creating a new strand S. In this way, double helical DNA can be copied precisely. (**d**) A circular superhelix DNA in which a helix is itself coiled into a helix

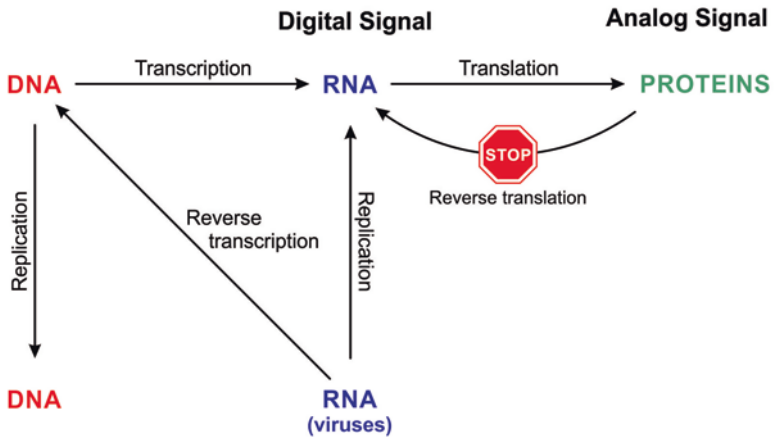


Fig. 16.4 Biological information transfer systems in DNA, showing digital and analog information carriers and transfer routes. Information flows from DNA to RNA to proteins, but it forbids reverse translation from proteins back to nucleic acids because the analog system cannot turn into a digital signal. The central dogma essentially captures the irreversibility of the translation process. DNA replication, transcription, and translation take place in all known organisms. RNA replication occurs in cells infected by some RNA viruses under the influence of enzymes encoded by the viral genome. The RNA of retroviruses is replicated by way of DNA with the help of a virally encoded reverse transcriptase

sequences of nucleic acids can determine the sequences of proteins. However, the sequences of amino acids in proteins cannot determine the sequences in DNA or RNA. The flow of information via sugar-phosphate backbone is like a one-way street.

With the discovery of reverse transcriptase (RT) by retroviruses, the transfer of information from RNA to DNA is possible, and the central dogma was modified. One fundamental rule in the central dogma still holds: the information flows from nucleic acids to proteins. Once the information has passed to proteins, it cannot get out again. In other words, the transfer of information from proteins to nucleic acids is impossible [18].

Koonin [19] offered a novel explanation for this unique, unidirectional route of information transfer in the central dogma. He suggested that nucleic acids carry one-dimensional, digital information, whereas folded proteins carry three-dimensional, analog information. The transfer of digital information from nucleic acids to the analog information embodied in folded proteins is permitted, but the reverse process of transferring information from proteins to nucleic acids is prohibited by the biological exclusion principle. The three-dimensional analog information format cannot be transferred to a one-dimensional digital format. The information inscribed in the three-dimensional structure of proteins is entirely contained in their primary structure, that is, in the sequence of their amino acid residue. The digital information from mRNA is lost in the protein folding and cannot be recovered. The most important argument in the central dogma is that it forbids reverse translation (Fig. 16.4).

16.7 The Immortal Coil

Watson and Crick's iconic DNA structure is probably the third revolution in biology, beginning with Darwin's natural selection, followed by Mendelian genetics, which has had a significant impact on both science and society. DNA has a mystique appeal in its form and function. DNA molecules are twisted in an elegant spiral of the double helix—the immortal coil—the blueprint of life (Fig. 14.6c). DNA contains the recipe for making proteins, which are the dominant worker molecules in cells. Before a cell reproduces itself, it must make a copy of its DNA by replication.

A detailed portrait of the DNA structure is revealed by imaging techniques such as X-ray crystallography, nuclear magnetic resonance, and super-resolution microscopy. DNA is a short, exquisite, and high-density storage medium for genetic information. DNA's information storage system is constantly being adapted to handle massive amounts of information. Scientists are harnessing the powers of recombinant DNA technology to transform science, medicine, genetic engineering, agriculture, and industry. Because everyone's DNA is unique—except for identical twins—it can be used to identify people, which is why forensic scientists collect samples of blood, saliva, hair, and like at crime scenes. In recent years, clustered regularly interspaced short palindromic repeats (CRISPR) technology has allowed us to edit DNA, giving us the potential to provide new treatments to some of the most feared diseases—not only cancer but also Alzheimer's disease, Parkinson's disease, and others.

DNA is the code of life, i.e., how every living organism on Earth stores its genetic information. When Watson and Crick discovered the DNA structure, they could not dream of a day when a genome of a 38,000-year-old Neanderthal fossil would be sequenced and revealed that we have inherited 5% of Neanderthal DNA in our genetic makeup as a parting gift. Neanderthals lived in western Eurasia among modern humans about half a million years ago, primarily split from the ancestors of *Homo sapiens* in Africa and then became extinct 38,000 years ago. Interbreeding between Neanderthals and modern humans led to an exchange of genetic makeup of each group. Both species were exposed to novel viruses that offered immunity against *Helicobacter pylori*—infectious bacteria that cause stomach inflammation and some types of stomach cancer—are believed to have been inherited from Neanderthals. A recent study has suggested that all regions of the world outside Africa have inherited a genomic region from Neanderthals associated with protection against coronavirus disease 2019 (COVID-19) viruses [20].

Two features of the DNA structure made a remarkable impact on biology for two reasons: its digital nature and its complementarity, where one strand of the double helix makes a perfect replication. DNA is the information-coded material of life and has the ability to completely transform one species to another by modifying itself in cells. The digital information of DNA underlies our ability to read truths about our identity, ancestry, and traits. DNA is our genetic autobiography. In almost every cell of our body, a personal copy of the DNA molecule can be found with all our genetic information. Although four nucleotides of DNA are the same from bacteria to humans, it is the order of nucleobases in which they are arranged that creates such an incredible variety of living beings. Regulatory networks of genes play a critical role in making organisms different. We are biological machines created by our DNA. In eukaryotes, DNA is mostly contained in their nucleus. In prokaryotes, DNA is a single circular chromosome, free-floating in the region of cytoplasm, called the nucleoid. Smaller pieces of circular DNA called plasmids may be present.

DNA has a remarkable property of replication that is crucial for life. It can replicate and copy its genes so that the information needed to build and maintain cells is passed down. The fidelity of replication is immense. A remarkable characteristic of the DNA information transfer mechanisms is their accuracy. The error rate of DNA replication, for example, is in the order of 10^{-9} or one mismatched molecule in one billion. DNA replication is also at the heart of reproduction. Replication and reproduction have gone on nonstop ever since the beginning of life, mediated by DNA molecules. In *The Selfish Gene,* Richard Dawkins coined the term 'immortal coils' for the double helix of DNA molecules. He suggested that in the beginning, a DNA molecule he calls the 'replicator' began to replicate itself and thus gain in number in the protected environment of a cell, a survival capsule [21]. With the change of environment, DNA molecules adapted to these changes for survival. So, it kept mutating until one version of itself helped the cell adapt to its new home and start dividing again. Since then, the highly adapted DNA molecules have survived, proliferated, diversified, and achieved immortality.

Perhaps, DNA was stabilized in LUCA by natural selection and began to replicate accurately. The double helix structure of DNA with four bases has remained immortal for the last four billion years. Still, the sequences representing the digital information have become unique for each species and have changed with biodiversity. The tree of life as an icon is elegant and powerful, conveying the essence of Darwinian evolution. The tree was based on the vertical inheritance of traits from parents to offspring across successive generations. Now we know that DNA is life's hereditary material, i.e., it holds and passes on the genetic information from parents to offspring. The information carried by DNA directs each organism's construction, maintenance, proper functioning, and reproduction.

From LUCA in the remote past, the digital information traveled from the root to three domains of life—bacteria, archaea, and eukaryotes—and then diversified over time by the successive division of boughs, limbs, branches, and twigs, and finally to the top of the tree. The tree of life embodies a tree of digital information. Therefore, DNA sequencing has helped scientists to fine-tune the topology of the tree of life.

16.8 Conclusions

The evolution of DNA from RNA was the penultimate milestone in the origin of life. DNA, like RNA, consists of long polynucleotide chains assembled from four nucleobases. There are two differences in the molecular structure. First, the sugar ribose in RNA is replaced in DNA by a slightly different sugar, deoxyribose. The second difference is the replacement of uracil by thymine, which is uracil with an added methyl group (CH_3). This change still allows thymine to pair with adenine. In contrast to RNA, almost all DNAs in nature are double-stranded, except in some rare viruses.

Viruses contributed major enzymes to the assembly of DNA on an RNA template in protocells. Thus, began the DNA world when DNA replaced RNA as the major genome of protocells. With the advent of DNA, replication of information was entirely dissociated from its expression. Because DNA is much more stable than mRNA with more storage capacity, it is an excellent archive for information systems in the form of base sequences. DNA progressively took over the replicative storage function of mRNA, leaving the latter for protein synthesis. Genetic information began to flow

from DNA to mRNA to proteins in a two step-process involving both transcription and translation.

DNA is transcribed to mRNA by RNA polymerases, and the mRNA is translated to proteins in ribosomes. From DNA to mRNA to proteins, the one-way flow of information is called the central dogma of molecular biology. DNA consists of a pair of nucleotide chains twisted together in an elegant spiral, i.e., a double helix or an immortal coil—both strands of double-stranded DNA store the same biological information in the form of a molecular code. The strands are held together by complementary base pairing. A gene is a length of DNA that codes for a specific protein, and most proteins act as enzymes that catalyze specific reactions in metabolic pathways. DNA replication would be central to the binary fission of the first cell.

References

1. De Duve C (2005) Singularities: landmarks on the pathways of life. Cambridge University Press, New York
2. Koonin EV, Senkevich TG, Dojla VV (2006) The ancient virus world and evolution of cells. Biol Dir 1:29. https://doi.org/10.1186/1745-6150-1-29
3. Forterre P (2006) The origin of viruses and their possible roles in major evolutionary transitions. Virus Res 117:5–16
4. Powner MW, Zheng SL, Szostak JW (2012) Multicomponent assembly of proposed DNA precursors in water. J Amer Chem Soc 134:13889–13895
5. Xu J, Green NJ, Gibard C, Krishnamurthy R, Sutherland JD (2019) Prebiotic phosphorylation of 2-thiouridine provides either nucleotides or DNA building blocks via photoreduction. Nature Chem 11:457–462
6. Gilbert W (1998) The RNA world. Nature 319:618
7. Harish A, Caetano-Anolles G (2011) Ribosomal history reveals origin of modern protein synthesis. PLoS One 7:e32776
8. Furukawa Y, Chikaraishi Y, Ohkuchi N, Ogawa NO, Glavin DP, Dworkin JP, Abe C, Nakamura T (2019) Extraterrestrial ribose and other sugars in primitive meteorites. Proc Nat Acad Sci 116:24440. https://doi.org/10.1073/pnas.1907169116
9. Chatterjee S (2018) The hydrothermal impact crater lakes: the cribles of life's origin. In: Kolb VM (ed) Handbook of astrobiology. CRC Press, Taylor & Francis, Boca Raton, pp 265–295
10. Koonin EV (2016) Viruses and mobiles elements as drivers of evolutionary transitions. Phil Trans R Soc B 371:20150442. https://doi.org/10.1098/rstb.2015.0442
11. Takeuchi N, Hogeweg P, Koonin EV (2011) On the origin of DNA genomes: Evolution and division of labor between template and catalyst in model replicator systems. PLoS Comp Biol 7(3):e1002024
12. Orlando L, Ginolhac A, Willerslev E (2013) Recalibrating Equus evolution using the genome sequence of an early Millde Pleistocene horse. Nature 499:74–78
13. Leu K, Obermayer B, Rajamani S, Gerland U, Chen IA (2011) The prebiotic evolutionary advantage of transferring genetic information RNA to DNA. Nucleic Acid Res 39:1–13. https://doi.org/10.1093/nar/gkr525
14. Freeman S (2005) Biological science, 2nd edn. Pearson Prentice Hall, Upper Saddle River
15. Watson JD, Crick FHC (1953) A structure for deoxyribose nucleic acid. Nature 171:737–738
16. Crick FHC (1968) The origin of genetic code. J Mol Biol 38:367–379
17. Forterre P (2005) The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. Biochemie 87:793–803
18. Crick FHC (1970) Central Dogma of molecular biology. Nature 227:561–563
19. Koonin EV (2015) Why the Central Dogma: on the nature of the great biological exclusion principle. Biol Dir 10:52. https://doi.org/10.1186/s13062-015-0084-3
20. Zeberg H, Pääbo S (2021) A genomic region associated with protection against severe COVID-19 is inherited from Neanderthals. Proc Nat Acad Sci 118(9):e20263092118
21. Dawkins R (1976) The selfish gene. Oxford University Press, New York

First Life

17

Every living cell, even the simplest bacterium, teems with molecular contraptions that will be the envy of a nanotechnologist.

—Alonso Ricardo and Jack W. Szostak, 2009

Events, contingencies, chemical evolution, energy, and information systems in hydrothermal crater vent environments led DNA protocells to a single population of first cells about four billion years ago. The conversion of DNA protocells to first cells was the crowning achievement, the last and greatest one in the origin of life. The ability to divide must have been a property of protocells from the moment of their existence. However, the mechanism of cell division, with DNA replication creating identical daughter cells, is a breakthrough innovation that defines the first life. The French biochemist and Nobel Laureate Francois Jacob put it poetically: 'The dream of every cell is to become two cells.' During cell division, all three information systems—analog (AIS), hybrid (HIS), and digital (DIS) —are transferred vertically from parent to daughter cells. Probably, the members of this population of first cells resembled present-day hyperthermophilic bacteria in their structure, ecology, and main metabolic pathways. The emergence of the first cells was the culmination of the extended prebiotic synthesis on Earth, perhaps for 500 million years after the birth of our planet. Binary fission was the turning point in Earth's history, an event horizon in the biosphere, transforming a rocky planet into *Gaia*—a living planet. Ever since that life and Earth have been coevolving in a continuing tango. When we compare our planet with our apparently lifeless neighbors, life's influence on Earth is more profound and pervasive than we ever suspected. Life itself, once it got started in our Goldilocks planet, kept it habitable for four billion years.

17.1 From Protocells to First Cells

The fossil record suggests that the first cells appeared on Earth almost four billion years ago. Morphologically, these Archean microfossils look like bacteria without any nucleus [1]. How the first cells arose from the DNA protocell and how these first cells began to reproduce, increase their population, adapt, and evolve is the central question of the origin of life. The ability to multiply in identical copies of itself is what makes life so utterly different from anything else in the known universe. As discussed earlier, the prebiotic chemical evolution from cosmic ingredients to a single-celled living organism probably took place in the hydrothermal crater vent environment in hierarchical complexity. These pioneer heat-loving microbes—hyperthermophiles—thrived around the volcanic vent of the crater that spewed several toxic gases such as CH_4, NH_3, SO_2, and H_2S, providing chemicals for nutrients and warmth. These hyperthermophiles in the vent environment were probably the first life forms on this planet [2].

Here, we reconstruct the first cells from the combined evidence of the prebiotic evolutionary history, the Archean microfossils, and the morphology and function of modern bacteria. From the fossil record of the oldest microbes, it appears that the first cells were probably simple spheres or rods. We reconstruct from biochemical pathways that the first cells were likely single-celled microbes with a minimalist design, but they acquired the essential cellular structure components for integration and reproduction. The plasma membrane was a simple, double-layered cell wall and lacked extensive, sophisticated, external, and internal membrane systems. Most likely, the outer membrane was leaky. It was probably filled with porin proteins, which formed small holes through the membrane, allowing passive diffusion. These holes were big enough for nutrients and ions to pass through passive diffusion but small enough to keep the cell's machinery inside. The membrane encased the cytoplasm in which all soluble components were packed inside. Like bacteria, the first cell might have contained only one circular chromosome in the cytoplasm—a large DNA molecule packed closely in the nucleoid. This chromosome was most likely longer than the diameter of the cell in which it was encapsulated, so the circular configuration of the

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_17

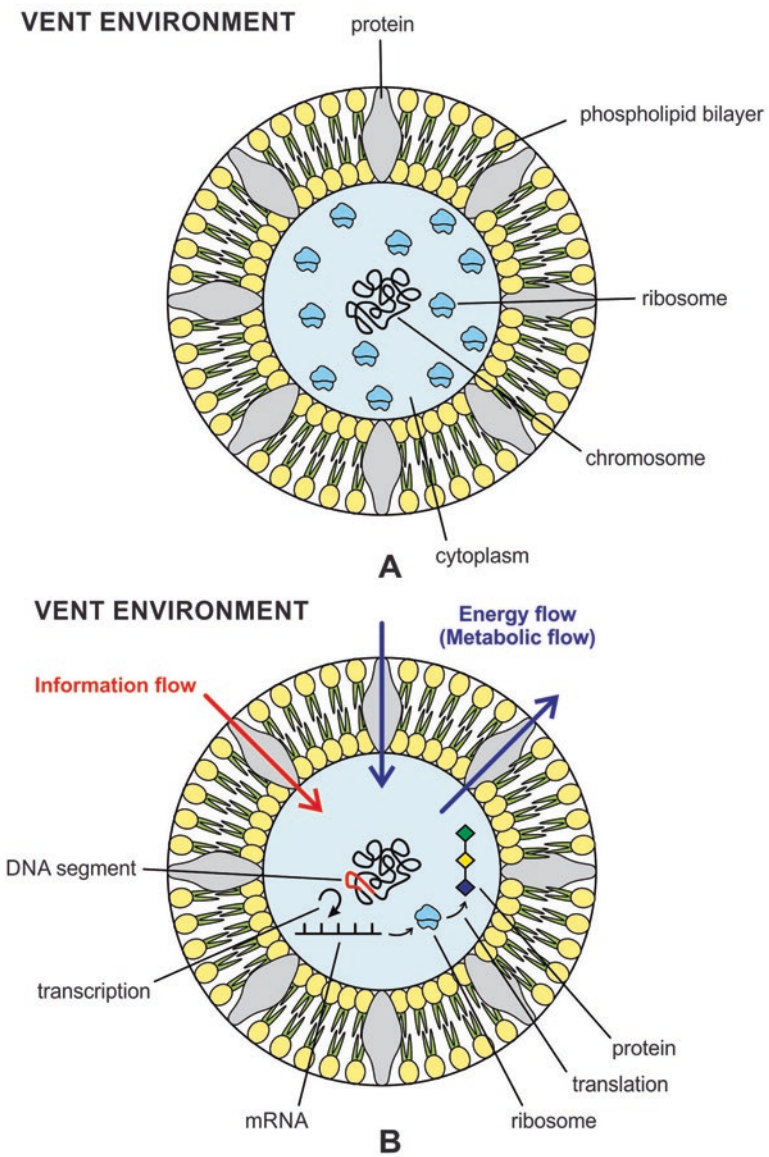


Fig. 17.1 (**a**) Reconstruction of the hypothetical first cell. The plasma membrane encloses a primitive cytoplasm that contains a prominent circular chromosome at the center; outside the chromosome, the most prominent structure was numerous ribosomes; other molecules like RNAs, various enzymes and proteins, and other translation machinery were too small to show. (**b**) Energy flow and information (or signaling) flow are represented as the basic exchanges of the first cell with the outer/inner environment. The vent environment's signal is transduced when the receptor proteins located in the plasma membrane attach to ligands on the exterior of the cell and then cause the behavior on the interior of the cell. The signal from the environment was received by the sugar-phosphate backbone of the specific gene to control gene regulation and expression: which genes to use and which not to use during the synthesis of proteins

chromosome was a space-saving device for the long chain of DNA. The DNA double helix coiled on itself, fitting into the nucleoid with the aid of enzymes to form a highly compact 'superhelix' circular structure (Fig. 16.3d), enclosed tightly in a condensed area of the nucleoid (Fig. 17.1a). Outside the chromosome, the most prominent structure in the soluble portion in the cytoplasm was dispersed ribosomes; other smaller molecules included various enzymes and other translation machinery for protein synthesis. Some molecules were devoted to producing energy from environments. The first cell was a bounded, autonomous entity with a plasma membrane that sealed it from outside environment, so that its internal cytoplasmic environment can be sequestered and self-contained to develop homeostasis. The cytoplasm contained chemical buffers so that acidity or alkalinity of the cell did not change, even when the chemical environment outside the cell changed. The first cell evolved to be autonomous, to survive as an independent living unit. Perhaps the flagella were not yet invented, as the fossil record suggests; the first cell lacked this propelling device to power movement.

It was known for many years before the DNA's detailed structure was determined that DNA is replicated during cellular reproduction. There is a sharp distinction between replication and reproduction. Molecules can replicate, but only cells can reproduce [3]. The first and most important attribute of life is reproduction. If something is alive ('autopoietic,' i.e., self-creating), then it can make a copy of itself (4). Reproduction with variation is an essential characteristic that distinguishes life from nonlife. This genetic variation is the raw material for Darwinian selection. The reproduction of the first cells was the watershed event in the history of life.

Without reproduction, life would have quickly come to an end, so reproduction is essential for the continuity and evolution of life.

The biological stage was the tipping point when the information stage converged into the biological stage. How do nonliving biopolymers such as membranes, nucleic acids, and proteins morph into living organisms—organizing into cells, reproducing, then growing, and evolving? The integrated and interconnected parts of the molecular systems of the first life produced homeostasis and discovered ways to respond to stresses, both physical and chemical. The major components of the first life were a plasma membrane, enclosing the core enzymes, and information carrier genomes, forming the integrated systems that included metabolizers and had the ability to capture energy from vent environments. The first life was autopoietic [4].

Building new proteins was one of the main tasks of the first cells. Over half of the molecules inside the *Escherichia coli* cell are involved in one way or another with the synthesis of proteins [5]. The recent estimate of *E. coli* genes has been 4401, of which 4285 genes encode proteins and 116 encode RNAs. The making of various types of proteins was one of the most important events for the first cell because a protein does more than just simply form structural components of the cell. It also composes thousands of enzymes that catalyze the production of the remaining biomolecules necessary for life. In early cells, different genes were active that produced only those proteins that were needed in the cell. All proteins in the first cell were encoded in one large circular DNA. The cell must control when and where each gene was used. The information held in the DNA genome was highly regulated. Perhaps a host of repressors and activators interacted with each gene, determining when it would be used to create proteins.

The first life orchestrated the control of gene expression at every step between the synthesis of messenger RNA (mRNA) and the activation of the gene product. In the first cell, the prebiotic information systems became increasingly sophisticated to process more and more advanced levels of biological information. In the peptide/RNA world, information flowed from mRNA to proteins [6]. With the emergence of DNA, the central dogma was established, which states that information flows from DNA to mRNA to proteins. With the emergence of the first cell, information flows from the environment to DNA to mRNA to proteins (Fig. 16.4b).

17.2 Gene Regulation and Signal Transduction

We can extrapolate from recent bacterial cells how first cells might have responded to an environmental cue. Protocells began to respond from environment initially with the development of the plasma membrane (Fig. 14.7e). First, cells needed to perceive and react to outside worlds to survive. They must have developed some mechanisms to transduce environmental information into gene expression. They could pick up external signals, which they then relayed to internal signaling pathways that directed their behavior. This surveillance could also trigger survival tactics for a variety of harsh situations such as lack of nutrients or drastic environmental changes. Signal transduction transfers a signal across a cell. Changes in gene expression allowed primitive cells to respond to environmental changes.

Bacterial cells can sense, remember, respond to, and adapt to their internal and external environments to maximize survival or accurately execute a development program. Cells have evolved complex regulatory and signaling systems capable of sophisticated information-processing tasks and have developed incipient memory for performing these tasks. Simple cells are open systems regarding both energy flow and information flow. An information transfer pathway is different from the metabolic pathway: the former is unidirectional, whereas the latter is bidirectional (Fig. 17.1b). In the information pathway or signal transduction pathway, only an impulse is related to the environment of the cell membrane.

First, cells were typically exposed to the ever-changing environment in which nutrient availability might have increased or decreased radically. They had signaled pathways that enabled them to detect and respond to changes in their surrounding environment. These microbes responded to such variations in their environment by altering their gene expression pattern. They expressed different enzymes depending on the nutrients available to them. Like protocells, first cells responded to environmental fluctuations of hydrothermal vents for survival and adaptation. For early cells to survive and reproduce in this environment, they must use resources efficiently, particularly resources that provide energy and nutrients. They developed a signal transduction pathway to relay messages from the environment to the cytoplasm of the cell for responses. Signal transduction converted an extracellular environmental signal into an intracellular signal. The extracellular signal can be light, temperature, nutrients, chemicals, toxins, osmotic pressure, or cell density, but intracellular signals are chemical messengers. Like bacteria, a two-component signal transduction pathway probably developed in the first cells for survival. First, the receptor protein receives a signal from the environment. Second, the response regulator protein, in turn, transmits the signal to the target. When ligands, signaling molecules, bind to receptor proteins, the receptor proteins change shape. The Receptor proteins located in the plasma membrane have binding sites for the signals they recognize. The scheme for receptor–ligand binding reveals the geometrical aspect to the matching and recognizing of these structures, an example of molecular shape recognition of analog

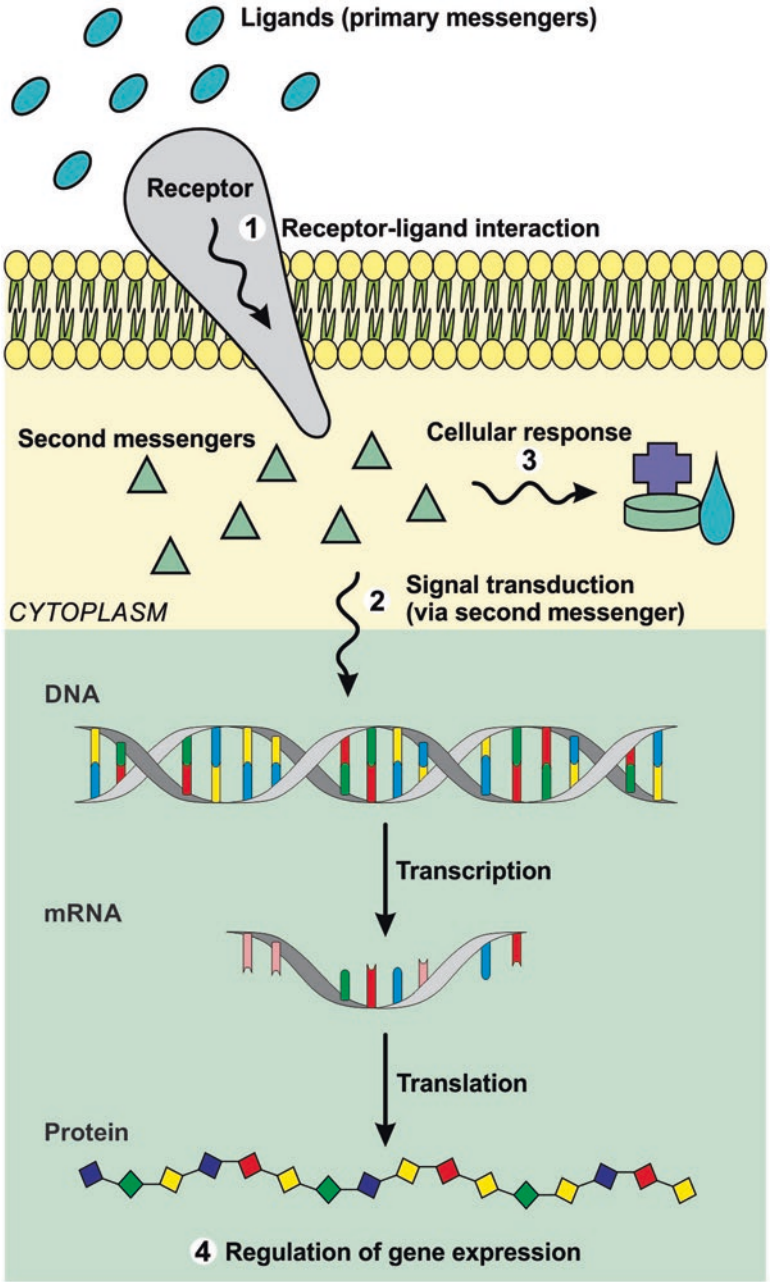


Fig. 17.2 Schematic representation of a signal transduction pathway in first cells. The signal is received from the environment when the ligand binds to the receptor. The signal is transduced when the receptor changes shape and becomes ready to cause a change inside the cell. The signal is amplified when the receptor causes a change inside the cell that activates molecules called the second messengers and is received by DNA for transcription and translation for protein synthesis

information systems. Ligand geometry is recognized by means of conformation of the binding site of the receptor. Spatial complementarity leads to the conformational fit.

The main features of signal transduction are shown in Fig. 17.2. First, cells responded to environmental changes by turning on genes for proteins that would help them survive. Genes were turned on and off by DNA-binding proteins that bound to DNA and controlled transcription.

The DNA sequence in a cell contains the information required to synthesize thousands of different proteins via mRNA molecules. Specific proteins must be synthesized in a timely manner to function properly in a cell. Genes are translated into proteins, and proteins instruct cell function. Thousands of genes are expressed in a certain way to determine what that cell can do. The transcription initiation is the primary control point for gene expression, which is usually at the beginning of the protein production process.

Bacteria have specific regulatory molecules that control whether a particular gene will be transcribed into mRNA. Often, these molecules act by binding to DNA near the gene and helping or blocking the transcription enzyme RNA polymerase. Genes are usually found in a cluster on a chromosome, where they are transcribed from one promoter (RNA polymerase-binding site) as a single unit. Such a cluster of

genes under the control of a single gene is known as an operon. Each operon contains regulatory DNA sequences, which bind regulatory proteins that promote or inhibit transcription [5].

First cells controlled and regulated the synthesis of proteins from information encoded in their DNA. Gene regulation is about early cells choosing which genes to use and which not to use for expression. The regulation of gene expression requires a significant amount of energy for an early cell to always express genes, so it is more energy-efficient to turn on the genes only when required. First cells must have developed a complete set of genes in their DNA molecules for blueprints of various kinds of proteins needed for survival and reproduction, yet these cells used only a fraction of those genes at any given moment. In general, a gene is expressed when its specific protein product is needed. First, cells might change what genes they were using depending on signals or changes in the environment. Transcription and translation occurred almost simultaneously in the cytoplasm of the first cells. Here, the gene expression was regulated at the transcription level (see Fig. 16.2a for the detailed transcription mechanism). Several mechanisms can regulate the transcription of a gene by an RNA polymerase. Protocells might have already developed constitutive genes for the essential housekeeping proteins that were typically in use all the time. First cells, in addition to constitutive genes, developed regulated genes, which were turned on and off as they were needed by the cells.

Gene expression occurs in two essential steps: transcription and translation. In bacteria, the two processes are tightly coupled in time and space and highly regulated in a fluctuating environment. In a bacterial cell, genes are organized into operons, or clusters of coregulated genes on the chromosome, where they are transcribed from one promoter (an RNA polymerase-binding site). These genes, close to the genomes, are regulated such that they are all turned on or off together when proteins are needed for a specific function. Other regulatory proteins are repressors and activators. Grouping related genes under a common control mechanism allow bacteria to adapt to changes in the external environment, rapidly optimizing the cell for survival at a given time.

In the population of first cells, most likely, gene expression was much simpler to control and modulate the single gene for one enzyme. In general, a gene was expressed only when its protein product was needed. Most likely, operons were gradually added as the first cell evolved and acquired more functional capability.

Five steps occurred during the flow of information in the first cell for gene expression, represented by arrows in the following expression (Fig. 17.2):

Environmental signal → DNA → mRNA → protein → activated protein.

The arrow from the environmental signal to DNA represents the selection of a gene by the cell to make a particular protein. The arrow from DNA to mRNA represents the transcription of that gene to mRNA. The arrow from mRNA to protein represents translation, in which ribosomes read the information in the mRNA and use that information to synthesize a protein. The arrow from protein to activated protein represents posttranslational modifications into a three-dimensional structure.

17.2.1 Quorum Sensing: Communication of Primitive Cells

Bacteria communicate using chemical signal molecules. They develop a robust capacity for memory that transmit from generation to generation. Many bacteria use a cell–cell communication system by gene regulation called *quorum sensing* in response to fluctuations of cell population density. In quorum-sensing systems, bacteria detect the density of other bacteria around them, creating microbial mats. An example of such group behavior is biofilm formation that requires coordinated chemical signaling between cells. Microbial mats are horizontally stratified, self-sustaining microbial communities, ranging in thickness from millimeters to several centimeters, formed by multiple biofilms of microorganisms embedded in a matrix of exopolysaccharides in a vertical manner due to physical gradients. Primitive cells—when enclosed in their protective, slimy biofilms formed with the assistance of quorum sensing—were able to withstand the extreme environment of hydrothermal vents. These biofilms also deterred viral infection, a smart protective shield. Quorum sensing uses signaling molecules known as autoinducers. These are hormone-like molecules continuously produced by bacteria, which can readily diffuse through the cell membrane. Microbial mats are complex communities of microbes, usually organized into layers. They were one of the pioneer ecosystems on Earth, together with stromatolites. Stromatolites are microbial mats in which sediments are trapped. Fossil stromatolites represent the earliest and most pervasive record of life on early Earth.

17.3 Cell Division

Life is organized into cells that grow and divide. The advent of cell division into two identical cells, each with a complete set of genetic machinery, defines the emergence of the first cells from their protocell precursors [8, 9]. Although protocells could divide by physical forces, this division was not controlled by DNA replication (see Fig. 9.4). The chromosomal replication was tightly linked to the first cell division. We can infer the origin of primordial cell reproduction from the modern bacterial fission and protocellular division.

Bacterial cells divide into two genetically identical cells during reproduction to propagate by binary fission. A typical bacterium, *Escherichia coli*, which colonizes the mammalian intestine, is an excellent exemplar of the cell cycle and binary fission. The cell cycle of a bacterium is the series of events in a cell that leads to the duplication of its DNA, the segregation of copied DNA, and the splitting of the parent cell into two identical daughter cells. A genome is a single, circular DNA chromosome that must be replicated and then distributed among the daughter cells; the cytoplasmic contents are also divided to give both new cells the machinery to sustain life. *Escherichia coli* can divide every 20 min [7].

Most bacterial cells spend their time cycling between a state of calm (interphase) and the dividing phase, known as the 'cell cycle.' The origin is close to the chromosome's binding site on the plasma membrane (Fig. 17.1). A cell grows twice its original size during interphase. This growth is coordinated with the duplication of the chromosome. Binary fission of a bacterial cell begins with DNA replication and segregation of the nucleoids. In bacterial fission, the circular DNA chromosome is attached to the cell wall near the cell's midpoint. DNA replication occurs bidirectionally as the cell grows and elongates. Before binary fission occurs, the cell must replicate its chromosome and segregate these copies to the opposite ends of the plasma membrane. Both cell division and DNA replication are coordinated so that the distribution of new DNA copies to each daughter cell is ensured. Several regulatory proteins orchestrate the function of bacterial growth and division. These core cycle regulators can cause key events, such as DNA replication or chromosomal separation, to take place. They also make sure that the cell cycle events take place in the right order. There is a coupling between DNA replication and cell division.

Bacteria reproduce asexually by binary fission, a process that produces identical offspring. A cell grows and doubles in size and then splits in half to produce two identical daughter cells. The first stage of cell division (or cytokinesis) is the formation of a Z-ring, composed of a tubulin-like protein, FtsZ, at the division site precisely at mid-cell. A constriction develops by the Z-ring between the duplicated chromosomes so that one is included in each of the two daughter cells. The fact that the bacterial chromosome is anchored to the plasma membrane makes this separation possible without the construction of an elaborate mitotic apparatus. This contractile Z-ring orchestrates the separation of the chromosomes and the cell division in the bacterium (Fig. 17.3). A septum is created between the nucleoids, extending gradually from the periphery to the center of the cell. The daughter cells separate when the new cell walls are in place, completing the cell division [5, 8].

As DNA is replicated, the replicated DNA molecules are segregated at the two ends of the cell in an energy-dependent process, forcing the cell to grow, elongate, and pull apart, making the equatorial region stretched, weak, and constricted. DNA replication exerts osmotic pressure on the cell wall by increasing the concentrations of entrapped biomolecules, and DNA separation actively drags the nucleotides apart. At this time, the cytoplasm is enriched with the FtsZ protein that aids the constriction of the membrane at the midpoint. The Z-ring establishes the division site's location, acts as the scaffold for the division apparatus, and provides the contractile force to precisely orchestrate the binary fission. As the contractile Z-ring closes like a purse string, the cell cytoplasm is divided into two, completing the cell division (Fig. 17.2). For the division to produce viable daughter cells, it must be coordinated in time and space with other significant cell cycle events, such as DNA replication and segregation. As the cell grows to twice its starting size during binary fission, the cytoplasm with essential hybrid and analog information systems is split into two identical parts, thus maintaining the same size and content of the cytoplasm of the parent cell [10, 11].

Cell division exhibits how digital, hybrid, and analog information systems are passed vertically from parent to daughter cells. Therefore, each daughter cell is self-sufficient with ingredients and information systems to jump-start its function. Eventually, each daughter cell grows to a regular size in the cell membrane and cytoplasmic content by taking resources from the environment and is ready for binary fission to increase the population and diversity [12].

The mechanism of cell division is a complicated multistep process. A new study shows that a mutant bacterium can reproduce without a wall or division machinery, supporting the idea that primordial cells could have divided using physical mechanisms, such as simple shearing alone [13]. Leaver et al. [14] generated a mutant strain of *Bacillus subtilis* that lacked cell walls. Although this mutant form contains a Z-ring, it does not participate in cell division. They found that these cells divide not by the Z-ring but rather by an extrusion resolution mechanism. This novel form of cell division provides insights into how early forms of cellular life may have increased. This pattern is strikingly like the 'pearling instability' seen in lipid vesicles. The study supports a model in which the constriction of the Z-rings is dependent on wall synthesis [8, 11].

Although a bacterial cell structure is extremely different from that of a DNA protocell because of billions of years of evolution, we may assume some common underlying mechanism of protocellular division from bacterial fission. The protocell might have achieved a rudimentary form of cell division using physical devices alone and was modified with the availability of proteins (see Fig. 9.4). Initially, DNA protocell division may have been a random burst when the cell size increased as the cell accommodated more complex biopolymers until it reached an unstable size. Surplus molecules were generated inside the protocell, causing it to bulge, and

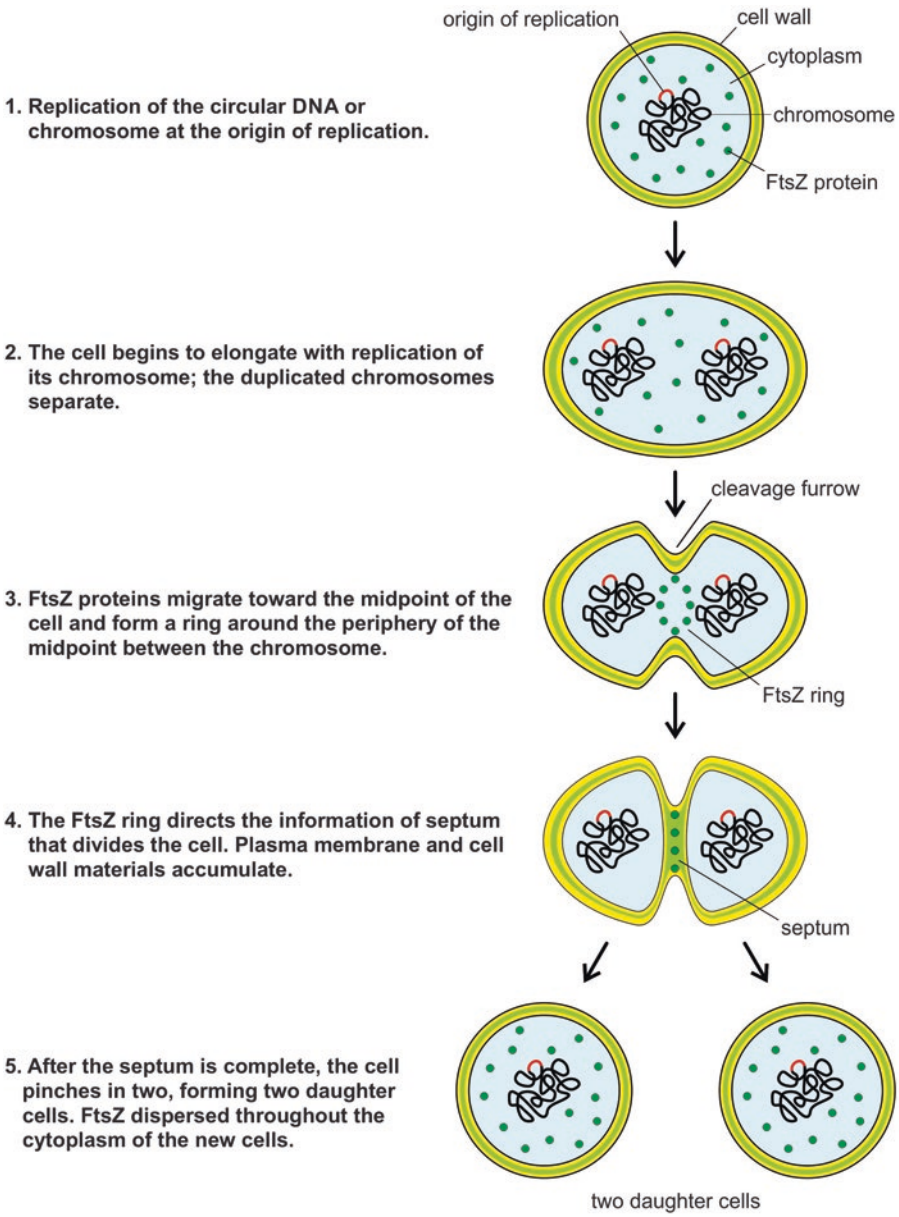


Fig. 17.3 A hypothetical scheme of cell division in the primitive first cell using a modern bacterium as a guide. The cell division cycle occurs in a cell, leading to duplication of its DNA. The early cell must coordinate its growth, division, cell volume, and shape with the genome's inheritance. During the process, thousands of FtsZ molecules come together in the middle of the cell and form a circle-like structure known as the Z-ring. Z-rings are produced in the middle of the cell division, leading to constriction. The replicated chromosome and cytoplasm separate into two new identical daughter cells

the protocell continued to elongate and became pinched in the middle. Eventually, the surface tension caused the cell to split into a pair of daughter cells. As the cell divided, the two new cells were not identical because there was no mechanism to ensure an equal distribution of the parent cell's contents. The dissimilarity in the daughter cells was advantageous at this stage of evolution because it promoted a diversity and variation upon which natural selection could operate. Still, actual cells required mechanisms that guaranteed identical daughter cells [15].

DNA replication was the primary driver of symmetrical binary fission in early cells [5]. However, it took many generations of DNA cells to perfect cell reproduction by countless trials and errors. We can speculate how the DNA cell might have learned how to reproduce identical daughter cells. Perhaps, initially, the cell grew, then elongated for cell instability, and pulled apart by physical forces, but divided asymmetrically without DNA replication, so one daughter cell might have intact DNA. At the same time, the other was devoid of it (Fig. 17.4). In the second generation of the attempt, the DNA might have replicated, but it was not coordinated with cell division, so one daughter cell might have two chromosomes. The other got only cytoplasm without any DNA. Eventually, the DNA protocell learned how to coordinate DNA replication with cell division, probably aided by few replicators.

The first cell was self-sustaining, DNA-based, and chemically sophisticated, possessing many housekeeping proteins and capable of mutation and Darwinian evolution. It had

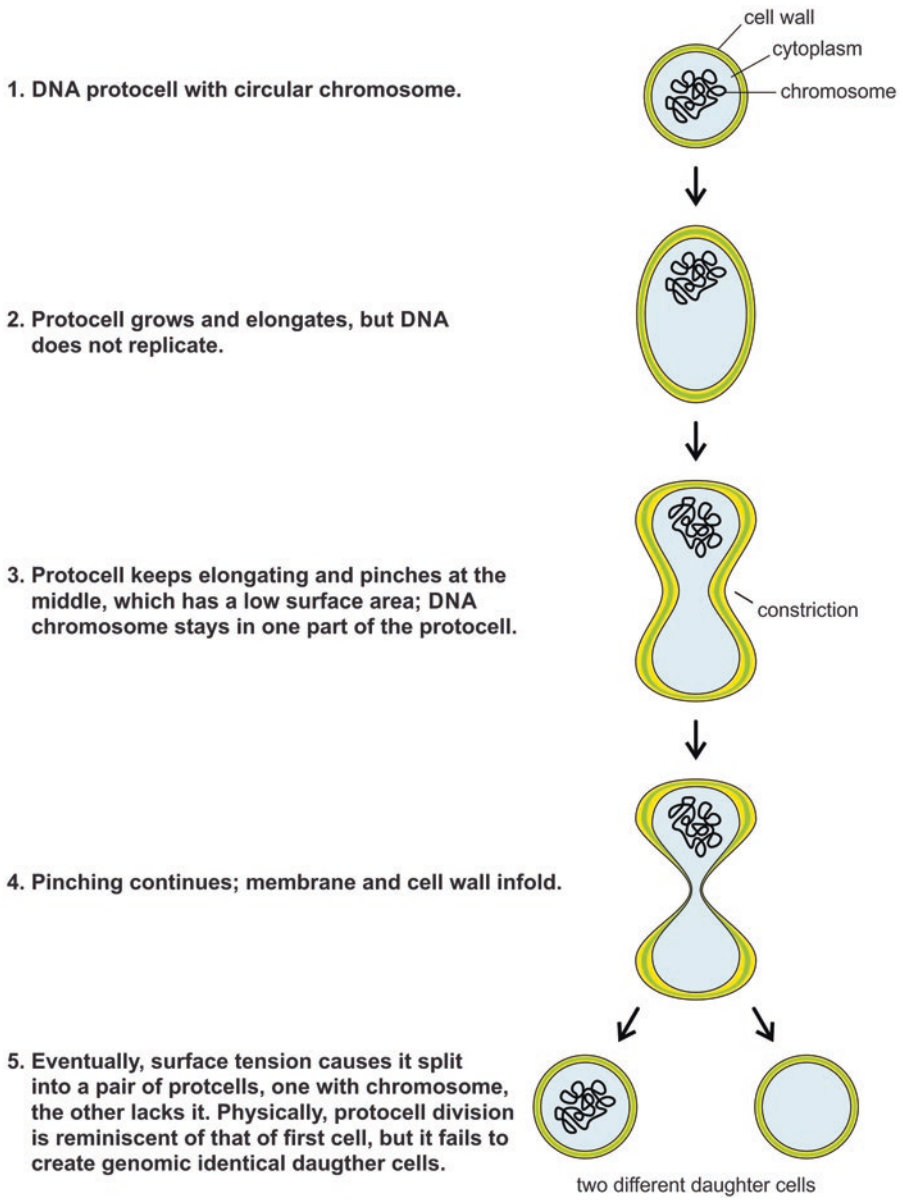


Fig. 17.4 DNA replication was the primary driver of symmetrical binary fission in early cells. Most likely, there were many trials and errors before the cell division was perfected. Here, a hypothetical scheme is shown where chromosomal duplication was not fine-tuned, giving rise to different daughter cells

developed capacities for harnessing energy from hydrothermal environments. However, these primordial organisms were continuously infected by viruses and eventually developed immune systems for their survival. The coevolution of viruses and the first cells was the source of innovation, gene enrichment, and diversity of early life. The first cells were stabilized, perfected in cell division, multiplied, and mutated for innumerable generations. They spread across hydrothermal systems on young Earth (Fig. 17.5).

The primitive cell doubled in size and split in half to produce two identical daughter cells. These daughter cells could then double in size again to produce four sibling cells and these to produce eight, and so on. When nutrients are plentiful, the doubling time of some bacterial species can be as short as 20 min. However, most bacterial species show a doubling time between 1 and 4 h. A single bacterial cell with a 1-h doubling time will produce more than one million offspring within 24 h.

17.4 Genetic Variation of First Cells

The first cells became the earliest self-sustaining organisms, and their emergence was the turning point in the early history of life. They quickly began to reproduce and multiply, colliding with each other and crowding the vent environment. Reproduction was an essential life process for the first cells because it allowed them to survive and continue as a population.

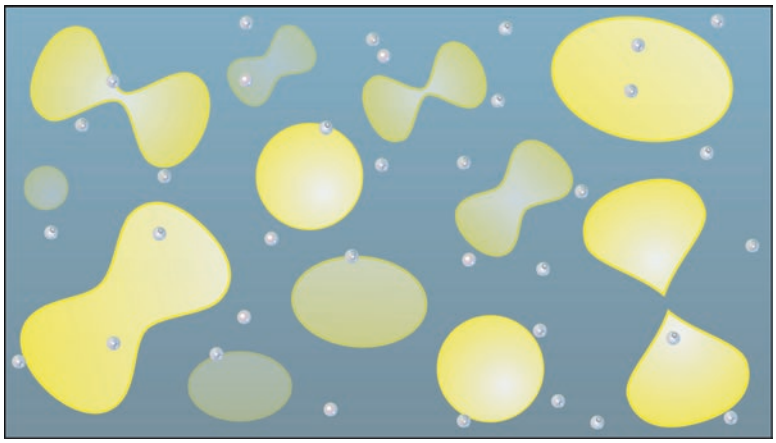


Fig. 17.5 Once the cell division was perfected, the hydrothermal vent environment was crowded with a new generation of daughter cells

This was the most momentous event in the early history of our planet, transforming the barren rocky planet into a living world.

Primitive cell populations, like modern bacteria, reproduced by splitting into two daughter cells via binary fission. Binary fission makes clones (genetically identical copies) of the parent cell. Since the daughter cells are genetically identical to the parent, binary fission does not provide an opportunity for genetic recombination or genetic diversity, aside from occasional random mutation. These primordial cell populations had little opportunity to survive unless they could adapt swiftly, using variation and diversity to escape extinction. Genetic variation is central to species survival, allowing groups to adapt to natural selection changes in their environment. A mutation is the source of new genetic variation because it changes the DNA sequences of an organism. Mutations are chance, haphazard events, more likely to be deleterious than useful. Harmful mutations are quickly eliminated from the population by natural selection. A beneficial mutation in an organism's DNA, on the other hand, gives the cell an advantage to cope with its environment and reproduce its kind slightly faster.

However, evolution by mutations alone would inevitably not only be chancy but also extremely slow and would be restricted to a single line of descent. Bacteria can be highly active in exchanging genetic information by horizontal gene transfer. Primarily, primitive cells exchanged genetic information between individuals—modern bacteria exchange genes through conjugation, transformation, and transduction mechanisms. In *conjugation*, DNA, such as a plasmid, is transferred from one cell to another via pilus, a hair-like protein tube found on the surface of bacteria. In *transformation*, a cell takes DNA from an environment, often free fragments of DNA that have been shed by other cells. In *transduction*, viruses that infect primitive cells move short pieces of chromosomal DNA from one cell to another by accident. Recurrent viral infections of protocells initiated the gene transfer, including DNA. Transduction seems to be a viable mode of gene transfer between cells. Such a gene transfer process between primitive cells allowing genetic recombination would enormously amplify and speed up the thrust toward diversity.

A combination of genes between two primitive cells was akin to primitive sex, but the new gene transferred horizontally rather than the vertical transmission of DNA from parents to offspring. Horizontal gene transfer (HGT) was a significant driver of variation of the primitive population of cells because the DNA was transferred from a successful lineage. HGT played an essential role in bacterial evolution. In primitive cells, the reproduction may have been high-speed. The short generation time, together with random mutations and genetic recombination, allowed the first cells to evolve quickly. Natural selection favored mutations and recombination that spread genes through a population of primitive cells. HGT helped the early cell population acquire new metabolic capabilities and adapt rapidly to environmental changes.

HGT continued in the primitive hyperthermophile cell population until the emergence of LUCA. Eventually, HGT created a new domain, archaea, when LUCA was split into bacteria and archaea. HGT between bacteria and archaea is still rampant, enriching the genomes of both domains. Eventually, in the Proterozoic, eukaryotes would split from archaea, creating three domains of life (Fig. 17.6).

Early cells faced another challenge: continuous assaults from viruses in the vent environment. Viruses were more abundant in the environment than were primitive cells. How did the primitive cells differentiate between 'foreign' and 'self'? How did they 'remember' a past encounter with viruses to produce a targeted response? One of the early innate immune systems by primitive cells was phagocytosis, a modified endocytosis process to ingest and destroy invading virus at the contact of the plasma membrane. These early cells might have developed immune systems like CRISPR to

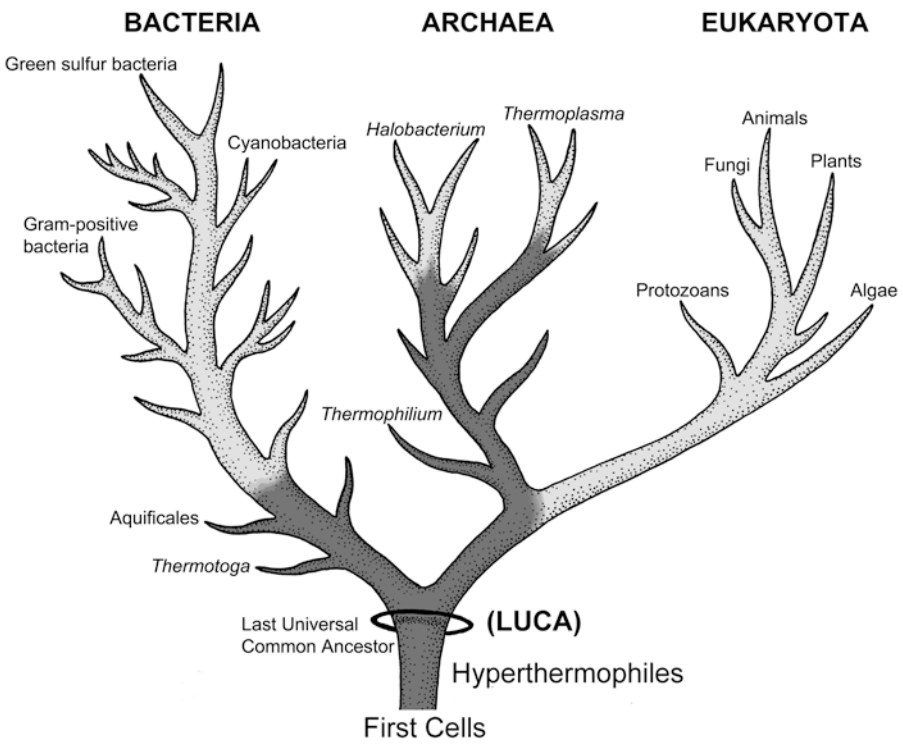


Fig. 17.6 The three domains of life—bacteria, archaea, and eukaryotes—based on comparisons of 16S ribosomal gene sequences in a living organism. Within the bacteria and archaea, the dark gray branches represent hyperthermophiles. LUCA is not the 'first' cell but is the 'last' before the divergence of bacteria and archaea. LUCA was a hyperthermophile bacterium-like organism and is estimated to have lived some four billion years ago (modified from Woese [18, 19] and Farmer [30]). The dark gray trunk region of the tree represents the hyperthermophilic lifestyle of early bacteria and archaea

recognize foreign viruses and protect themselves. Primitive cells likely developed a CRISPR-like immune system for survival. For billions of years, the planet's life comprised entirely of microbes such as bacteria and their viruses, the bacteriophages. Phages became a planet force and remain so even today, determining the makeup of the oceans and atmosphere, among other things.

HGT was the primary mechanism for the spread of viral resistance in primitive cells. Few early resistant cells acquired mutations from a viral infection; they passed the resistant gene to their descendants. A transformation that conferred resistance to viruses would give early cells a reproductive advantage to prevent viral attacks. There was another arsenal available for early cells. One of the most effective strategies would be to cooperate so that resources and protective measures are shared in the populations. They lived in colonies and produced biofilms by quorum sensing, a glue that clusters the population. A biofilm was a protective shield that could withstand assaults from viruses. The oldest stromatolites preserved the bacterial mat, suggesting that a biofilm was an effective deterrent from a viral infection of primitive cells.

17.5 LUCA: The Genetic Portrait of the Ancestor of Life

Darwin's theory of common descent with modification is the central pillar of modern evolutionary biology [17]. It states that all life on Earth, which has ever lived, has descended from one original primordial form that diverged with time, like tree branches from a single trunk. Darwin recognized that species not only evolve but also divide. As species evolve, they split and diversify through time, increasing morphological divergence like the branching pattern of a tree. He presented a metaphorical tree in his book 'On the Origin of Species' that showed how species change through time from the common ancestor. Darwin's tree is a visual representation that shows how a common descent relates to all species.

Darwin left it to later biologists to figure out the topology of the real evolutionary tree. Before developing sequence-based molecular methods, it was impossible to know the evolutionary relationships connecting all of life and thereby draw a universal evolutionary tree. Although the topology of the tree of life has changed over time as more and more genetic and proteinic information about organisms becomes available, they all confirm Darwin's theory of common descent. A genetic portrait of the ancestor of all living things has been slowly emerging in recent years. This venerable ancestor, the last universal common ancestor or LUCA, was most likely a single-cell, bacterium-like organism that lived in the hydrothermal vent environment.

We know all cells are monophyletic, descendants of a single founder cell, rather than having polyphyletic or multiple, independent origins. This view is supported by a list of circumstantial evidence: the unity of the genetic code, the universal use of the same biomolecules, such as DNA, RNA, and protein, and the homochirality of certain key molecular structures, such as L-amino acids and D-sugars,

from bacteria to humans. Today's organisms contain numerous proteins that are clearly homologous, most notably ATP synthase, an enzyme that creates the energy storage molecule ATP. Therefore, the organisms themselves must be homologous, descended from a common ancestor, the most recent of which we understand as LUCA. All extant life is genetically related, providing a unifying foundation for the web of life.

Darwin's dream of the tree of life was realized on the grandest scale when Carl Woese [18, 19] proposed that all cellular life can be placed in three separate fundamental domains—bacteria, archaea, and eukaryotes—based upon sequence comparisons of ribosomal RNA (rRNA) sequences (Fig. 17.6). According to the 'three-domain tree,' the eukaryotes and archaea are closer to each other than they are to bacteria. The 'three-domain tree' is the most visible image depicting the diversity of cellular life, but it has not gone unchallenged. An alternative' two-domain tree,' in which eukaryotes are nested inside archaea, has gathered support from recent phylogenetic analyses [20].

Molecular phylogenetics has provided overwhelming evidence that all living organisms on Earth descended from a single ancestral form, the last universal common ancestor (LUCA). LUCA's concept is central to studying the early evolution of life's origin, yet its nature and phylogenetic position in the tree of life is controversial. Woese [19] identified LUCA as 'progenote,' an organizational level like an RNA protocell that preceded the first cell. A similar view of LUCA has been suggested by Forterre [21]. Today, LUCA is considered a sophisticated organism with a complex DNA-based structure recognizable as an actual cell. LUCA is probably a bacterium-like organism that could help establish how early life began on Earth. LUCA is not the 'first' cell but is the 'last' before the divergence of the two-domain tree—bacteria and archaea. Before LUCA, there was life even when the first cell was mutating and evolving into diverse populations [22]. LUCA represents the most recent common ancestor of all extant organisms.

Irrespective of whether life on Earth had an ultimate common ancestor is a subtle question, complicated by the phenomenon of horizontal gene transfer (HGT). Present evidence suggests that HGT is a widespread phenomenon in the bacterial world. Some scientists believe that LUCA was not an organism but a collection of diverse organisms exchanging their genes by rampant HGT without constraint; LUCA, in this view, looks like a tangled tree of life contradicting its monophyletic origin [23]. The same objection applies to the possibility that specific genes that were not present in LUCA arose later, in separate lines, by convergent evolution. We have enough information to sketch a portrait of the universal ancestor with due regard to these uncertainties.

Theobald [24] analyzed the vast array of molecular sequences now available from the three domains of life using robust statistical techniques and concluded that LUCA had a monophyletic origin, regardless of HGT or multiple origins of life. He studied amino acid sequences from 23 universally conserved proteins in the 3 domains of life. He then applied standard programs for inferring evolutionary trees. His study was based on several simple assumptions about how the diversity of proteins arose. He applied a model of selection theory to a molecular phylogeny that favored the existence of a single origin of LUCA over an extensive suite of alternative hypotheses. According to Theobald, LUCA was a microbe living in the early Archean world from which all life evolved.

Weiss et al. [25] genetically analyzed 6.1 million protein-coding genes and 286,524 protein-coding clusters from sequenced prokaryotic genomes representing various phylogenetic trees; he then identified 355 protein families that were probably common to LUCA. The team realized that HGT between bacteria and archaea about four billion years ago masked much of LUCA's original genetic signal. Genes found in both bacteria and archaea could have been shared through HGT and would not necessarily have originated from LUCA. The team searched for 'ancient' genes with exceptionally long lineages, but which did not seem to have been shared around by HGT on the assumption that these ancient genes came from LUCA. Once they finished their analysis, they found only 355 genes that belonged to LUCA and can tell us something about how LUCA lived. LUCA shares these genes with two groups of modern microbes: *Clostridiaum*, a genus of thermophilic bacteria, and the methanogens, a group of hydrogen-metabolizing archaea. Most likely, LUCA lived in an anaerobic hydrothermal vent, rich in H_2, CO_2, and iron. LUCA was chemosynthetic and autotrophic, deriving free energy in the form of redox potentials and pH gradients from the vents. Furthermore, hyperthermophiles have adapted to vent conditions, so maybe life began here.

Although the physiology and habitat of LUCA in the hydrothermal vent environment, as suggested by Weiss et al. [25], have been endorsed by previous researchers, their conclusion about the portrait of LUCA—a progenote, only 'half-alive'—has created a great deal of controversy. Of course, such a small number of genes (~355) would not support life as we know it, and critics immediately seized on its apparent gene shortage, pointing out that essential components capable of nucleotide and amino acid biosynthesis were missing. Gogarten and Deamer [26] have rightly criticized that these authors reintroduced an old misconception of LUCA, a progenote that most scientists rejected. LUCA had evolved far beyond the origin of life; it was a full-fledged prokaryote cell that could accomplish the complicated task of synthesizing proteins.

LUCA was a sophisticated organism preceded by a long period of Darwinian evolution. The emergence of the first cells and that of LUCA are separate events; the former led to the latter. As we discussed earlier, there was life before LUCA. Once LUCA appeared, it quickly gave rise to two domains—bacteria and archaea (Fig. 17.6). Although LUCA is long gone, its closest relatives may still be with us. LUCA itself was a hyperthermophilic chemoautotroph, a view supported by molecular phylogeny [16, 26–29]. Allowing for uncertainty, LUCA was not significantly different from modern hyperthermophilic organisms that appeared 4–3.5 Ga before the split between bacteria and archaea [30].

17.6 Archaea: Emergence of a New Domain of Life

DNA sequence comparison and structural and biochemical comparisons suggest that LUCA split into two domains in early Archean: bacteria and archaea [18–20]. HGT was the primary driver of the divergence of archaea from bacteria. Archaea constitute a domain of single-celled prokaryotic organisms. They are diverse, ecologically important, single-celled microorganisms but are more evolved than bacteria in their genetic sequences. They have unique properties, such as methanogenesis and cell envelope composition, although many characters are shared with bacteria through ancestry or HGT. The members of archaea and bacteria are united in the informal group prokaryotes by similar general cell size and shape, the lack of a nuclear membrane and organelles, and the presence of a large circular chromosome occasionally accompanied by one or more circular DNA plasmids. Despite their morphological similarity, archaea possess genes and several metabolic pathways more closely related to eukaryotes. Like bacteria, archaea also lived in hydrothermal vent environments during the early evolution of life. Most likely, bacteria diverged from lineage-producing archaea and eukaryotes, giving rise to three domains of life (Fig. 17.6). Eukaryotes are more derived from the other two domains of life, archaea, and bacteria, and appeared much later in the fossil record, around 2.1 billion years ago.

Because of their morphological similarity and similar paleoecology during the early origin of life, the shape of microbial fossils does not usually differentiate between bacteria and archaea. However, chemofossils may help discriminate their identity. Archaeal membranes do not contain the same lipids that bacteria do. Instead, their membranes are formed from isoprene chains. These chains are unique to archaea, and they make useful markers for the presence of ancient archaea. They also left chemical fingerprints behind as methanogens (methane producers).

Most of the HGT genes in archaea originate from bacteria; the remaining genes are transferred either from other lineages of archaea or, in rare cases, from eukaryotes. The exchange of genes by HGT is a major driving force behind genome evolution in early microbial life. The geological setting of the earliest microbial life is generally considered a hot hydrothermal crater vent environment, supporting the physiology and habitat of LUCA, bacteria, and archaea [2, 20, 27, 28].

17.7 Conclusions

The first cells probably arose on Earth about four billion years ago. They became autonomous to maintain internal homeostasis for survival an independent life. One of the important events leading to the first cell must have been the ability to reproduce by duplicating genomes and then dividing the parent cell into two identical daughter cells. These primordial cells were probably similar to the living hyperthermophilic bacteria. These early cells were plasma membrane-bound structures with highly organized, dynamic interiors. DNA replication was central to the binary fission of the first cells. Like binary fission of living bacteria, the first cells grew, duplicated all major cellular constituents, like plasma membrane, cytoplasm, RNA, DNA, ribosomes, etc., distributed this content, and then divided into two nearly identical daughter cells. With the onset of binary fission, the population of primitive cells grew rapidly in the hydrothermal vent environment, undergoing Darwinian evolution and diversification by mutation and horizontal gene transfer. The important point is that variation of daughter cells could occur during cell division. The altered base sequence then produced the variation of the population of cells essential for evolution by natural selection, which is another unique property of life.

The hypothesis that life originated in a scorching environment is supported by present-day hyperthermophiles, the direct descendants of LUCA. LUCA is not the ‘first’ cell but is the ‘last’ before the divergence of bacteria and archaea.

References

1. Schopf JW (1999) Cradle of life: the discovery of Earth’s earliest fossils. Princeton University Press, Princeton
2. Chatterjee S (2018) The hydrothermal impact crater lakes: the cradles of life’s origin. In: Kolb VM (ed) Handbook of astrobiology. CRC Press; Taylor & Francis, Boca Raton, pp 265–295
3. Dyson F (2004) Origins of life. Cambridge University Press, Cambridge
4. Maturana HR, Valera FJ (1981) Autopoiesis and cognition. The realization of living. Boston Studies in the Philosophy of Science, Reidel Publishing, Boston
5. Freeman S (2005) Biological science, 2nd edn. Pearson Prentice Hall, Upper Saddle River
6. Chatterjee S, Yadav S (2019) The origin of prebiotic information system in the peptide/RNA world: a simulation model of the

evolution of translation and the genetic code. Life 9:25. https://doi.org/10.3390/life9010025
7. Goodsell DS (2010) The machinery of life. Springer, New York
8. Chen IA (2009) Cell division: breaking up is easy to do. Curr Biol 19:R327–R328
9. Deamer DW (2011) First life: discovering the connections between stars, cells, and how life began. University of California Press, Berkeley
10. Zwicker D, Seyboldt R, Weber CA, Hyman AA, Jülicher F (2017) Growth and division of active droplets provides a model for protocells. Nature Phys 13:408–413
11. Saha R, Verbanic S, Chen IA (2018) Lipid vesicles chaperone an encapsulated RNA aptamer. Nature Comm 9. https://doi.org/10.1038/s41467-018-04738-8
12. Deamer DW, Pashley RM (1989) Amphiphilic components of the Murchison carbonaceous chondrite: Surface properties and membrane forming formation. Orig Life Evol Biosph 19:21–38
13. Kurihara K, Okura Y, Matsuo M, Toyota T, Suzuki K, Sugawara T (2015) A recursive vesicle-based model protocell with a primitive model cycle. Nature Comm. https://doi.org/10.1038/ncomms9352
14. Leaver M, Domingez-Cuevas P, Coxhead JM, Daniel RA, Erington J (2009) Life without a wall or division machine in *bacillus subtilis*. Nature 457:849–853
15. Panno J (2010) The cell: the nature's first cell life form. Facts on File, New York
16. Stetter KO (2006) Hyperthermophiles in the history of life. Phil Trans R Soc 361:1837–1843
17. Darwin C (1859) On the origin of species. John Murray, London
18. Woese CR (1994) The universal ancestor. Proc Natl Acad Sci U S A 99:6864–6859
19. Woese CR (2002) On the evolution of cells. Proc Natl Acad Sci U S A 99:8742–8747
20. Williams TA, Foster PG, Cox CJ, Embley TM (2013) An archaeal origin of eukaryotes supports only two primary domains. Nature 504:231–236
21. Forterre P (2006) The origin of viruses and their possible roles in major evolutionary transitions. Virus Res 117:5–16
22. Cornish-Bowden A, Cardenas ML (2017) Life before LUCA. Jour Theor Biol 434:68–74
23. Doolittle WF (2000) Uprooting the tree of life. Scient Amer 282(2):90–95
24. Theobald DL (2010) A formal test of the theory of universal common ancestry. Nature 465:219–223
25. Weiss MC, Sousa FL, Mrnjavic N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF (2016) The physiology and habitat of the last universal common ancestor. Nature Microbiol. https://doi.org/10.1038/NMICROBIOL,2016.116
26. Gogarten JP, Deamer D (2016) Is LUCA a thermophilic progenote? Nature Microbiol. https://doi.org/10.1038/NMICROBIOL2016229
27. Reysenbach AL, Shock E (2002) Merging genomes with geochemistry in hydrothermal ecosystems. Science 296:1077–1083
28. McCollom TM, Shock EL (1997) Geochemical constraints on chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal systems. Geochim Cosmochim Acta 61:4375–4391
29. Dodd MS, Papineau D, Grenne T, Slack JF, Rittner M, Prajno F, O'Neil JO, Little CTS (2017) Evidence for early life in Earth's oldest hydrothermal vent precipitates. Nature 543:60–65
30. Farmer JD (2000) Hydrothermal systems: doorways to early biosphere evolution. GSA Today 10:1–9

The Habitat and Nature of Archean Life

18

The earliest well-preserved sedimentary and volcanic rocks are Archean and provide insights into atmospheric composition, climate, and life.

—David C. Catling and Kevin J. Zahnle, 2020

18.1 The Hadean Environment and Life

We have outlined the possible pathways for the origin of life from cosmic ingredients in hydrothermal crater vent environments to the first cells. Our final quest is when life first emerged, which has been an enduring and frustrating mystery. The answer may come from the fossil evidence of the earliest known life forms. However, the fossil record of early life is extremely fragmented and deteriorates further back in time to the Paleoarchean (3.6–3.2 Ga) and Eoarchean (4–3.6 Ga) eras. Moreover, we have no rock records from the Hadean Eon. Because of plate tectonics and constant recycling of Earth's crust, only a handful of rock outcrops remain that are older than 3.5 Ga. These rocks are metamorphosed and highly deformed by heat and pressure, making it extremely difficult to detect unequivocal signs of life. The problem with the early fossil record of life is that it is limited in occurrence, controversial, and difficult to interpret.

However, fossils are not the only evidence of early life. Another line of evidence for the origin of life exists in the genomes of two modern domains of life, bacteria and archaea, which evolved during the Archean. By comparing their genomes, we can infer the age of LUCA, the common ancestor of bacteria and archaea. A molecular clock model suggests that LUCA may have existed prior to the late heavy bombardment (LHB) (>4.1 Ga), during the most violent period in our planet's history soon after the moon-forming impact [1]. The estimated age provides the crude upper (older) age limit of the origin of life further back in time in the Hadean Eon and is significantly earlier than the oldest fossil evidence would suggest.

The genomic evidence is indirectly supported by the discovery of the detrital zircon grains as old as 4.4 Ga from Australia [2] and 4.2 Ga from India [3], the oldest surviving primordial crustal material. The oxygen isotopic ratios in detrital zircons indicate the presence of liquid water on Earth's surface in the Hadean eon [4].

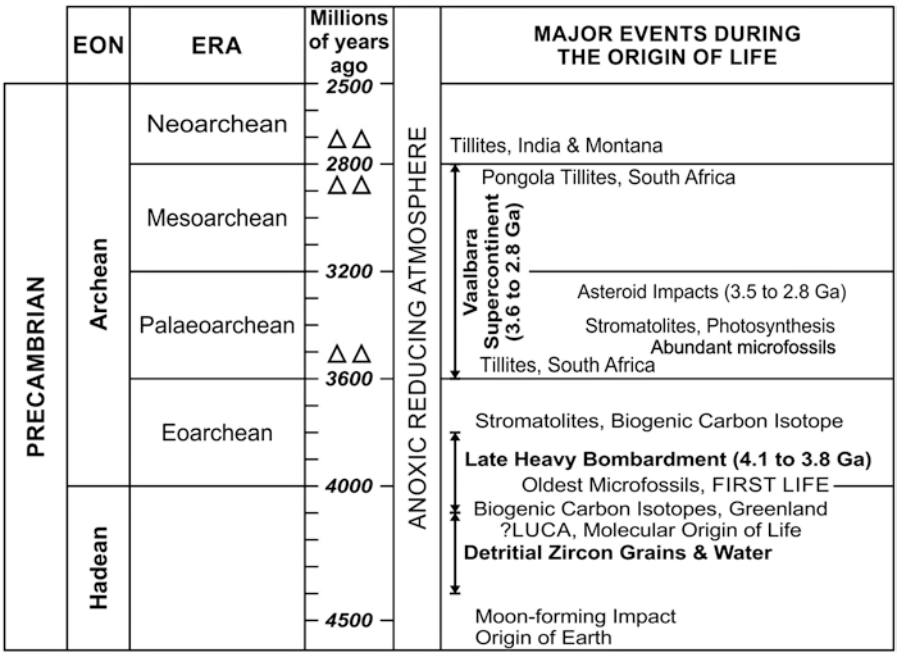
On the other hand, the direct evidence of life's origin has come from the fossil record in the Eoarchean sediments, which provides the lower (younger) age limit of the origin of life. So far, the oldest fossil evidence, in the form of filaments and tubes, has come from the jasper-carbonate banded iron formations of hydrothermal vent environment sediments from 4.3 to 3.8 Ga from the Nuvvuagittuq Craton of Quebec, Canada [5, 6]. Since there is no rock record of Hadean, we put the upper limit of the oldest microfossil age at the Hadean–Eoarchean transition. The biological origin of some putative fossils has received criticism because some abiotic chemical processes can display similar morphologies [7, 8]. In our view, the authors of these findings presented conclusive evidence from the Nuvvuagittuq greenstone facies as the oldest, undisputable record of life beyond the scope of pseudofossils.

The Earth today is extremely different from the Eoarchean world when life began around four billion years ago. It was a watery planet where the global ocean stretched from pole to pole. Small continental islands were few and probably clustered in groups showing highly cratered surfaces as likely cradles for abiogenesis. Archean volcano-sedimentary rocks in these ancient protocontinents are potential sites for biosignatures such as chemofossil, microfossil, and stromatolite records. Continental to shallow-water hydrothermal habitats in five Archean greenstone belts has yielded the oldest fossil remains. These are the Nuvvuagittuq Craton of Canada, the Isua Craton of Greenland, the Kaapvaal Craton of South Africa, the Pilbara Craton of Australia, and the Singhbhum Craton of India (Fig. 6.1). We describe the biosignatures from these cratons in the subsequent sections.

The Eoarchean surface environment surrounded by the global ocean was quite harsh with the presence of active volcanism and asteroidal impacts during the late heavy bombardment (LHB). Because of active volcanism, asteroidal impacts, and the low rate of organic burial, the temperature

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_18

Table 18.1 The integrated timescale for the origin of life during the Hadean-Archean eons, showing major geological, paleontological, and genomic events. (Modified from [9])



of seawater was significantly higher (between 55 and 85 °C) than it is today [8]. In summary, at the end of the Hadean, Earth had oceans, continents, and an anoxic atmosphere that was likely rich in CO_2, N_2, and perhaps life [9]. There was no free oxygen in the atmosphere or ocean water.

In our view, impact bombardments would have affected highly reducing environment of the Hadean. Perhaps, the Hadean represents the age of abiogenesis, assuming the long processes of prebiotic synthesis, but the origin of the first cells might have occurred at the Hadean–Archean transition. Here, we propose a timescale for the origin of life by integrating geological, genomic, and fossil evidence (Table 18.1).

18.2 Eoarchean and Paleoarchean Biosignatures

Although detrital zircons from Australia and India suggest that Earth's continental crust existed as early as 4.4 Ga, the oldest crusts are exclusively Eoarchean cratons, composed of tonalite–trondhjemite–granodiorite (TTG) series with enclaves of greenstone packages (basalt, komatiite, and hydrothermal sedimentary strata) (see Sect. 6.2.1). The Acasta Gneiss in northwestern Canada (~4.03 Ga) from the Hadean–Archean transition is the oldest known crust. Eoarchean rocks of TTG–greenstone assemblages are known from the Nuvvuagittuq Craton of Canada, the Isua/Akilia Craton of Greenland, the Kaapvaal Craton in South Africa, the Pilbara Craton in Australia, and the Singhbhum Craton of India (see Fig. 6.1). The volcano-sedimentary rocks of the greenstone facies of Eoarchean and Paleoarchean terranes host all the known types of biosignatures: chemofossils, microfossils of primitive cells and microbial mats, and stromatolites. Generally, these biosignatures of early life are preserved in hydrothermal cherts, formed by chemical precipitation of silica-rich gels, which encapsulate the delicate structure of early cells and organic matter. The hydrothermal settings for the early life range from terrestrial crater lakes with impact spherule layers to riverine environments, volcanogenic edifices, and deposits such as pillow lavas to shallow-water sandstones and carbonates in a pre-plate tectonic regime [10]. A handful of Paleoarchean terranes have survived relatively unscathed despite their ancient age. The earliest indirect evidence of life on Earth dates back to the Late Hadean (~4.1 Ga) in chemofossils in the form of biogenic carbon [11]; the microfossil record extends from ~3.8 to 4.3 Ga [5, 6] (see Table 18.1).

Earth has been habitable for at least four billion years. The earliest rock record indicates the presence of a microbial biosphere from 4 to 3.8 billion years ago. The microfossils from the Nuvvuagittuq Craton of Canada are tiny filaments and tubes in a hydrothermal vent setting, which provided a habitat for Earth's hyperthermophilic earliest life-forms [5, 6]. Although the recognition of ancient microfossils by their morphology alone is difficult, there are chemical traces of life within rocks from Archean hydrothermal vents [7, 8, 12–15].

The record of Archean microfossils is sparse, but we can infer the origin and affinities of the earliest life forms from the molecular phylogeny of recent microbial communities. It is generally believed that LUCA, at the base of the tree of life, was a simple prokaryote with features between those of

bacteria and archaea, which lived under extreme conditions like modern hyperthermophilic microbes, such as *Thermotogales* and *Methanococcales* [15–18]. LUCA represents a phylogenetic event horizon, a point in the history of life beyond which no phylogenetic analysis is possible. Most estimates suggest that LUCA was a hyperthermophilic microbe with the smallest genome and the fewest genes of any free-living organism, i.e., just 500 genes, the minimum essentials required to survive and reproduce [19]. Today, hyperthermophilic bacteria and archaea grow optimally above 80 °C, with an upper limit of growth of up to 113 °C. They form complex communities in hydrothermal systems in terrestrial and submarine environments. Hyperthermophiles thriving in high-temperature terrestrial hydrothermal craters have stimulated new theories of life's origins [20, 21].

Impact-generated hydrothermal systems have been present throughout Earth's geological history, and ancient Archean hydrothermal deposits provide clues to Earth's earliest biosphere [21, 22]. The hydrothermal systems of old greenstone belts, including the Nuvvuagittuq Craton of Canada, the Akilia/Isua Craton of West Greenland, the Pilbara Craton of Western Australia, the Kaapvaal Craton of South Africa, and the Singhbhum Craton of India, are potential sites for finding the earliest traces of life.

Our planet has encrypted its history in the physical and chemical features of sedimentary rocks. Part of this history is paleontological. Paleontology is the most crucial tool for studying the emergence of life in the oldest volcano-sedimentary rocks—the Eoarchean greenstone belts. The biosignatures of earliest life fall under three categories:

1. Carbonaceous remains of microbial cells (chemofossils)
2. Bona fide cellular fossils (microfossils)
3. Microbially influenced organo-sedimentary structures (stromatolites)

Burial is an integral part of fossilization because walls and envelopes are more likely to survive microbial decay because of mineral encapsulation. Thin layers of hydrothermal cherts would have preserved microorganisms from the overlying water levels, whereas mineral precipitation could fossilize microorganisms in three-dimensional detail in the Eoarchean and the Paleoarchean. The fossilized microbial remains, such as cells, colonies, biofilms, mats, and generic degraded organic matter, were sealed and better preserved because of encapsulation by silica during deposition. On early Earth, hydrothermal activity was widespread in the Archean terranes, and microbial mats provided a suitable substrate for mineral precipitation [10, 19]. Early primitive cells inhabited these hydrothermal environments and left their footprints.

It is generally believed that microbial life could not have established a permanent toehold on young Earth until the Late Heavy Bombardment subsided ~3.8 Ga [23]. However, microbes are tough and resilient and can survive in a hostile world. They can live in conditions that would kill other life forms, including intense cold, heat pressure, dehydration, acidity/alkalinity, radiation, and other chemical and physical extremes. Bacteria, transformed into dormant spores, can survive millions of years in extreme environments. The microbial occupation of Hadean and Eoarchean Earth (~4.1 Ga) in the Jack Hills of Australia, the Nuvvuagittuq Craton of Canada (4.3–3.8 Ga), and the Akilia and Isua supergroups of Greenland (~3.8 Ga) suggest that the habitable zone was not completely abolished on Earth during the terminal stage of the Late Heavy Bombardment [24–26]. Perhaps the subsurface habitats of hydrothermal impact crater lakes provided a sanctuary for the origin and early evolution of life [20].

18.2.1 Chemofossils: Indirect Evidence for Life

The organic matter of microorganisms in an anaerobic environment would be preserved but would undergo gradual degradation with the loss of H, O, N, and functional groups. With continued metamorphism in the Archean greenstone belt, the carbon-rich molecules turned into kerogen and finally to graphite. Sulfur components from microbes may also be preserved. Distinctive chemical signatures of organic matter may record these degraded fragments of the ancient microbes. Chemofossils indicate the first traces of life on Earth in the Late Hadean from detrital zircon grains (~4.1 Ga) (Table 18.1). The indirect record of early life in the sedimentary rocks from Archean greenstone belts is in carbon and sulfur isotopes, which are chemical fingerprints or chemofossils [27, 28]. Metabolic processes produce distinctive isotopic fractionations when specific carbon and sulfur isotopic forms are selected from the chemical or organic substrates accessible to organisms. The recognizable $\delta^{13}C$ and $\delta^{34}S$ signatures in sediments that occur after burial in both residual organic matter and associated minerals are particularly useful tools for understanding the nature and extent of microbial activity. Indirect evidence suggests that life existed on Earth around four billion years ago in the form of chemofossils.

In living organisms, carbon-12 (^{12}C) dominates the heavier isotope, carbon-13 (^{13}C), whereas, in inorganic limestones, carbon-13 is dominant. The sulfur isotopic ratio provides another biosignature of early life. The earliest indications of life on Earth are biogenic carbon (carbon-12) isotopic measurements preserved in a 4.1-billion-year-old zircon from the Jack Hills, Western Australia [11]. This discovery suggests that life appeared during the Late Heavy

Bombardment period and survived it. The finding in the Jack Hills indicates that the early Earth certainly was not a dry, boiling, hellish planet but became more conducive to life around 4.1 Ga. The graphite inclusions in this crack-free and isotopically undisturbed zircon crystal from the Jack Hills are interpreted as the biogenic remnants of early microbes, not inorganic fragments.

The next record of biogenic carbon (carbon-12 or ^{12}C) is from the early Archean Akilia and Isua supracrustal rocks of Greenland, which are about 300 million years younger than the carbon inclusions from the Jack Hills [24–28]. Unfortunately, these greenstone rocks in Greenland have been subjected to intense metamorphism so that any microfossils, if they ever existed, have been destroyed by heat and pressure. The Greenland graphitic residues are enriched in carbon-12, which probably indicates that they are chemical traces of early microbial life. However, carbon isotopes of this type cannot tell us which organisms were present in Greenland.

18.2.2 Microfossils

Microfossils are preserved remains of microbial organisms. Morphological, geochemical, and isotopic data imply that life was relatively widespread and advanced in the early Archean period, with metabolic pathways analogous to those of recent prokaryotic organisms. These Archean microfossils are microscopic, only a few tens of microns in size, and need an electron microscope to study. They are preserved in the forms of spheroids, ellipsoids, rods, or filaments, thanks to the presence of walls or extracellular envelopes that resisted postmortem decay. However, similarities to extant microorganisms should not be taken as positive evidence for biogenicity because abiotic processes can produce various morphologies [7, 8].

The record of Archean microfossils is rare. Few authentic fossil assemblages were preserved in shallow-water hydrothermal settings, and they are typically associated with laminated, stromatolitic sedimentary rocks. Microfossils from submarine hydrothermal systems have not been reported in Precambrian rocks, although hyperthermophilic microbes are ubiquitous in modern seafloor hydrothermal settings [29, 30]. In the next section, we will discuss the microfossils from the Archean volcano-sedimentary rocks of the greenstone.

18.2.3 Stromatolites

Stromatolites provide important indirect evidence on the distribution of microbial mat communities. Microbial mat would deter viral infection of early cells and provide insights into the dynamics between viruses and their hosts. Stromatolites are macroscopic, laminated organo-sedimentary structures accreted from microbial growth, movement, or metabolism. The shape of a stromatolite is mainly governed by the setting in which it builds. The flat-layered ones indicate that the environment was quiet, perhaps a lake or a shallow lagoon. The convex-upward flexures, forming domes or columns, indicate tidal flats. Stromatolites are trace fossils of microbial activity and thus provide weaker evidence of life than do microfossils.

18.3 The Earliest Records of Life on Earth

18.3.1 Microfossils from the Nuvvuagittuq Craton of Canada

Based on the carbon dating technique, the oldest putative microfossils, at least 3.8 billion years and possibly as old as 4.3 billion years, have been reported from the greenstone facies of the Nuvvuagittuq Craton on the eastern shore of Hudson Bay, Canada. The diverse microbial communities occurred in the form of delicate tiny filaments and tubes in jasper-carbonate banded iron formations (BIF) in the submarine hydrothermal vent environment and provided a habitat for Earth's first life forms reminiscent of hyperthermophilic bacteria [5, 6, 29, 32]. The BIF occurs as orange-to-red outcrops interlayered with metavolcanics. The authors emphasized that modern iron bacteria produce similar structures and grow as filaments around hydrothermal vents, subsisting on iron chemicals and constructing tube-shaped holes in the sediment. Similar filaments filled with iron compounds have been found in the Nuvvuagittuq rocks. These fossils are surrounded by minerals containing biogenic phosphorous. The rocks have yielded organic carbon (carbon-12) that bacteria have created. Carbon isotopic values of carbonates and carbonaceous matter are consistent with the biogenic origin of these microbial communities [33–35]. Perhaps, at least 3.8 Ga or even earlier, a diverse microbial ecosystem was established on primordial Earth. Recognizing ancient microfossils just by morphology turned out to be difficult. Since abiotic processes can produce similar structures, some researchers have disputed the biotic origin of these filaments [7, 8]. The authors [5, 6] provided multiple lines of evidence to confirm biogenicity assessment and refute any abiotic origin of these microfossils. If proven, the groundbreaking discovery will surpass the previously confirmed 3.5–3.7-Ga microfossils. However, the lack of Haden rocks suggests that life might have emerged around early Eoarchean.

18.3.2 Stromatolites from the Isua Craton of Greenland

The Isua rocks of Greenland have yielded a unique biogenic carbon signature from the greenstone facies (~3.8 Ga)

[26–28], but it was controversial: irrespective of whether the putative chemofossil had been created by ancient life forms or represented artifacts. Recently, Nutman et al. have described stromatolites of possible bacterial origin from the 3.7-billion-year-old Isua greenstone belt in Greenland [30]. The stromatolitic structures are about 1–4-cm high and have an overall conical shape with low amplitude and an internal finely layered texture. They are compositionally different from interlayered bedded sedimentary structures, but they lack organic or cellular remains. Cyanobacterial communities that produced these stromatolites suggest that life had invented photosynthesis and harvested solar power as early as ~3.7 Ga. With the inventions of photosynthesis, cyanobacterial life may have invaded shallow seas and distributed globally. The presence of cyanobacterial stromatolites from the Isua greenstone belt, however, pushes the origin of hyperthermophilic life back further, to near the start of Earth's sedimentary record. Both the discovery of biogenic carbon and stromatolites from the Isua greenstone facies is consistent with the recent discovery of hyperthermophilic microfossils from the Nuvvuagittuq rocks of Canada [5, 6] and the molecular genetic clock that placed life's origin in the Hadean Eon (>4 Ga) [1, 31]. The Isua greenstone belt may be another site where early hyperthermophilic life prospered, which gave rise to photosynthetic bacteria.

18.3.3 The Vaalbara Supercontinent of Australia and South Africa

The close stratigraphic correlation between the Onverwacht Group of the Kaapvaal Craton in South Africa and the Warrawoona Group of the Pilbara Craton in Western Australia suggests that they were once part of the ancient supercontinent Vaalbara (see Fig. 6.7). They have yielded the oldest and best-preserved early Archean microfossils from the hydrothermal volcano-sedimentary rocks. These microfossils provide critical evidence for the habitat and nature of early life on Earth in the ancient Vaalbara continent. The stratigraphic correlation of the Onverwacht Group with the Warrawoona Group suggests that South Africa and Western Australia were connected to form the oldest supercontinent Vaalbara. It formed at the beginning of 3.6 Ga and remained intact until it was broken apart by ~2.8 Ga [33, 34]. The sequences of the Onverwacht and Warrawoona Groups consist of several kilometers-thick impact-generated basaltic to komatiitic lava flows, which are interbedded with sediments; these sediments were silicified to chert by hydrothermal activity. Fossil microorganisms from these chert horizons have yielded single-celled bacterial and archaeal microorganisms that were already thriving ~3.5 Ga. The Vaalbara supercontinent contains stratigraphic evidence for four significant bolide impacts in both the Pilbara and Kaapvaal Cratons between 3.5 and 3.2 Ga, revealed from four horizons of spherule/ejecta layers when life had diversified [34].

It appears that early life survived catastrophic bombardments in terrestrial hydrothermal vent environments that have important implications for looking for the earliest forms of life on Mars and elsewhere in the solar system [34, 35]. Hydrothermal cherts interlayered with pillow lavas (~3.5 Ga) are a significant component of Archean greenstone belts. The presence of microbial remains on silicified crater floors in hydrothermal vent environments appears to be favorable habitats for Archean life [20, 21, 38, 39]. In the following sections, we separately discuss the fossil records from the Pilbara and the Kaapvaal Cratons, which both show abundant microfossils in the early Paleoarchean (≤3.5 Ga) hydrothermal sediments. This is the crucial time when microbial communities became widespread and diversified.

18.3.3.1 Microfossils and Stromatolites from the Pilbara Carton, Australia

The Pilbara Craton of Western Australia represents one of the great geological storehouses hosting the early evolution of life on Earth [37]. The Pilbara Supergroup (~30-km-thick) contains one of the famous Archean volcano-sedimentary rocks for searching for early evidence of microfossils. The Warrawoona Group (~3.5–3.4 Ga) is the oldest sequence; it contains interlayered thin chert horizons and felsic volcanic rocks; the Kelly Group overlies it. Three sequences within the Warrawoona volcaniclastic rocks may host evidence for Earth's earliest life. From the oldest to the youngest, these formations represent the Dresser Formation (~3.49 Ga) at the bottom, the Apex Chert (~3.46 Ga) in the middle, and the Strelley Pool Formation (~3.43 Ga) at the top. Hyperthermophilic bacteria and archaea were the major components of the ancient microbial activity, as evidenced by carbonaceous remains and fragmentary remains of cell walls from the hydrothermal black cherts of the Warrawoona. The fossil record suggests that two domains of life were already split from the LUCA during this time. The chemofossils from the Jack Hills of Western Australia reinforce the view that life originated on Earth at least 600 million years earlier in the Hadean Eon [11].

In the North Pole area, the Dresser Formation (~3.49 Ga) consists of a complex volcano-sedimentary unit containing the earliest record of stromatolites; however, their biogenicity has been debated for over three decades. Chert-barite beds of the Dresser Formation contain chemofossils of biogenic carbons. Hyperthermophilic microfossils in methanogenic vent environments have been found in bedded chert and hydrothermal black silica veins in the Dresser Formation [38]. Both unbranched and tubular types of filaments are found in the bedded chert, indicating the activity of methanogens or methane-producing hyperthermophilic archaea. Recently, Djokic et al. have discovered a variety of microbial activity and stromatolites from the Dresser Formation in terrestrial hot spring deposits containing geyserite [39]. It is a mineral deposit formed from silica-rich fluids with near-boiling

temperature, which occur exclusively in terrestrial hot springs and hydrothermal craters. This finding shows that a diverse variety of life existed in the terrestrial environment very early in Earth's history.

Noffke et al. [40] identified various morphotypes of stromatolites from the Dresser Formation, including domal, nodular, wavy-laminated, conical, and stratiform, which thrived in flat coastal environments. They correlated these microbial mats with anoxygenic photosynthetic bacteria that produced these stromatolites. In that case, complex photosynthetic communities appeared quickly through mutation from hyperthermophiles during the early radiation of life. This finding suggests that microbial life invented anoxygenic photosynthesis from H_2S earlier than previously believed.

The microfossils from a chert unit within the Apex Basalt Formation (~3.46 Ga) of the Pilbara Craton have generated a great deal of controversy. The microbial assemblage of the Apex chert consists of 11 taxa of filamentous microorganisms, which are classified as cyanobacteria [12]. However, the biogenicity of the Apex microfossils has been questioned [13] but defended [41]. I believe that the Apex microbial fossils are among the oldest evidence of bacterial remains, but a cyanobacterial affinity cannot be confirmed solely by the morphology. Given that the filamentous microfossils are embedded in hydrothermal silica, barite, and native metals, typical of the hydrothermal environment, it seems likely that the Apex microbiota were ancient hyperthermophilic bacteria and archaea [42].

The Strelley Pool Formation (~3.43) has yielded rich assemblages of both stromatolites and microfossils and has been accepted by many researchers as the oldest, undisputable record of life [1]. The formation is formed in a shallow-water environment and is sandwiched between the Warrawoona and Kelly Groups. The stromatolites' complex morphology occurred in a silicified carbonate unit, which was most likely biologically mediated [37].

The discovery of filamentous microfossils of sulfur-based bacteria from the Sulfur Springs Group (~3.2 Ga) suggests the diversification of thermophilic bacterial life in hydrothermal vent environments [37, 43]. The fossils are well-preserved, showing cell-like structures. The cells are clustered in groups and associated with pyrite grains. These microfossils possibly represent anoxygenic photosynthetic bacteria like our modern sulfur bacteria. They used the bacteriochlorophyll pigment to harness sunlight and manufactured sulfur and glucose as byproducts. Pyrite crystals associated with the microfossils are highly likely to be the byproducts. The sulfur bacterium has been considered an important transition during the evolution of cyanobacteria.

18.3.3.2 Stromatolites and Microfossils from the Kaapvaal Craton, South Africa

The Kaapvaal Craton of South Africa is another extraordinary storehouse for the earliest record of life, such as primitive hyperthermophilic bacteria and archaea from the hydrothermal cherts sandwiched between pillow lavas [35]. The Barberton greenstone belt of the Kaapvaal Craton represents the oldest known volcano-sedimentary sequences that have provided critical evidence of early life. The Swaziland Supergroup, about 23-km-thick, of the eastern Transvaal represents one of the oldest, unmetamorphosed volcano-sedimentary sequences. The Swaziland Supergroup is divided into three distinct units: the basal volcano-sedimentary Onverwacht Group, the middle sedimentary Fig Tree Group, and the top the Moodie Group, all containing rare archives of early life. The Onverwacht Group (~3.5 Ga) is mainly composed of volcaniclastic chert alternating with basalt layers, like the sequences of the Warrawoona Group. Both groups have yielded four distinct layers of impact spherules and are valuable niches for Earth's oldest fossils [34].

Both paleontological and geochemical data have suggested the existence of a rich microbial ecosystem, possibly containing hyperthermophilic bacteria and methanogenic archaea as early as 3.8 Ga [21, 38, 44–46]. In this environment, microbes lived in the cracks and crevices of impact glasses of melt rocks in hydrothermal crater vent environments, harnessing energy and nutrients from the glass by dissolving it. The microbial community is represented by filamentous and spherical structures of carbonaceous matter and microbial mats in cherts [37, 46–49]. Westall et al. [36] recognized the bacterial fossils from the finely laminated Onverwacht cherts (~3.3–3.5 Ga) as small spherical and rod-shaped structures of a biogenic origin. Walsh et al. [49, 50] reported long, filamentous microfossils from the Onverwacht Group. In the 3.5-billion-year-old lava from the Barberton greenstone belt in South Africa, Banerjee et al. [47] discovered microtubules bored by hyperthermophiles that resemble modern microbe borings. The margins of the tube contain organic light carbon, suggesting that microbial life colonized these underwater volcanic rocks around hydrothermal vents. This biologically mediated corrosion of synthetic glass is a well-known phenomenon in the microbial world [35]. The habitat of microbial life in the Onverwacht Group is probably analogous to modern black smokers or hydrothermal crater lakes.

Tice and Lowe [51] discovered anoxygenic photosynthetic microbial mats from the Buck Reef Chert (~3.4 Ga) of the upper Onverwacht Group; these mats are preserved in chert in shallow marine environments. These microfossils show filamentous structures, and the organic matter

preserved in these rocks appears to be of biological, not hydrothermal, origin. The bacteria that inhabited Earth at that time were anaerobic photosynthetic bacteria (such as purple and green bacteria) existing in a primitive ecosystem devoid of molecular oxygen.

The overlying Fig Tree Group consists of terrigenous clastic sedimentary units interstratified with volcaniclastic and volcanic rocks. It has yielded stromatolites, kerogens, and spheroidal microfossils from the chert bed [34]. Knoll and his mentor Barghoorn [52] documented a well-preserved population of spheroidal carbonaceous microfossils that contain isolated individuals and commonly paired cells and the intermediate stages of binary fission. Microbial life from the 3.2-billion-year-old Moodie Group in the form of bacterial cells associated with stromatolites in tidal-flat settings represents another remarkable discovery of the early diversification of life.

Noffke and a coworker [40] reported microbial mats of anoxygenic bacteria from the siliciclastic deposits of the Moodie Group (~3.2 Ga) associated with stromatolites. These stromatolites show carpet-like, laminated fabric typical of microbial mats. This discovery suggests that anoxygenic photosynthetic bacteria thrived in siliciclastic tidal-flat settings. The microbial mats of anoxygenic bacteria created the sedimentary structures in the tidal flats of the Moodie Group. Similarly, Javaux et al. reported the oldest and largest Archean, organic-walled spheroidal microfossils from the siliciclastic deposits of the Moodie Group [53]. These large organic-walled microfossils resemble cyanobacteria.

In summary, the fossil record of the Pilbara Craton and the Kaapvaal Craton of the ancient Vaalbara supercontinent suggests that early life initially emerged as hyperthermophilic bacteria and archaea, then evolved to anoxygenic photosynthesizers, and finally to oxygenic cyanobacteria. We have recreated the Vaalbara ecology by showing a cross section of three crater basins that flow from terrestrial to shallow marine settings and are interconnected by underground networks; this represents the record of Earth's first microbial life during the Archean (Fig. 17.6).

18.3.4 Microfossils and Stromatolites from the Iron Ore Group of the Singhbhum Craton of India

Algoma-type banded iron formations (BIFs) from the Iron Ore Group (IOG) of the Singhbhum Craton of eastern India have yielded precise 3.4 Ga age from zircons [54]. These are thinly bedded, chemical sedimentary rocks comprising alternating layers of iron-rich minerals and chert. These rocks are typically intercalated volcano-sedimentary sequences that formed in terrestrial hydrothermal vent settings within greenstone belts [55]. They differ from superior-type BIFs, India's largest iron ore deposits, representing extensive units, mainly Proterozoic in age, deposited in a shallow marine environment and showing no specific association with volcanic units.

The IOG represents the most spectacular hydrothermal deposits of a volcano-sedimentary sequence in a terrestrial setting. It begins with lower basaltic andesitic lava (~ 2 km) at the base, followed successively by lower shale (2.3 km) and chert and tuffaceous rock (30 m), banded iron ore (0.7 km), upper shale with manganese ore, and upper lava at the top [54]. The volcano-sedimentary sequence with Algoma-type BIF of the IOG is highly similar to the Onverwacht Group of Kaapvaal Craton of South Africa and the Warrawoona Group of Pilbara Craton of Western Australia.

Maithy et al. [56] reported spheroidal and filamentous microfossils from the lower shale of the IOG. These fossils are encased in the hydrothermal chert matrix. The preserved structures contain carbonaceous or iron-stained three-dimensionally preserved forms. The biogenicity of the microfossils is evident from the presence of a wide laminated sheath around individual cells, the presence of dark content within the cells indicating cytoplasmic shrinkage or vacuole, and binary fission of the cells. These microfossils are associated with spectacular stromatolites. These fossils provide potential evidence that ancient life thrived in India's IOG, a region composed of 3.4-billion-year-old hydrothermal crater lakes, the old cradle that nurtured early life. Perhaps these microbes represent hyperthermophilic bacteria (filamentous forms) and archaea (spheroidal forms). The biosignatures of the IOG are like those of the Dresser Formation of Australia of similar age.

18.4 Evolution of the Archean Biosphere

The first record of microfossils and chemofossils on Earth dates to the early Archean or even Hadean during the Late Heavy Bombardment, in which the first cellular life and LUCA originated in a harsh environment [21]. Because the photic zone was vulnerable to significant asteroidal impacts, life must have developed in the sheltered benthic environment such as a hydrothermal crater lake. Molecular phylogenetics suggests that LUCA was hyperthermophilic [15]. Deposition environments of greenstone facies hosting early microfossils, biogenic carbon, and stromatolites are quite diverse, ranging from terrestrial hydrothermal craters and coastal hydrothermal fields to submarine hydrothermal vents, consistent with the notion of a hyperthermophilic origin of life [19, 21].

According to the best available evidence, bacteria-like hyperthermophilic organisms were already present on Earth at least 3.8 billion years ago or even in the Hadean [5, 6].

Around 3.5 Ga, life diversified into two distinct domains, bacteria and archaea. Soon after, there was a dramatic change in the metabolism of microbes. Instead of harnessing thermal energy from the vent, they began to explore a new energy source from the Sun, when anoxygenic photosynthetic bacteria and then oxygenic photosynthetic bacteria appeared quickly. These early forms of single-celled microbes were dominant life forms during most of the Archean. LUCA (which had more than 500 genes) evolved to cyanobacteria (containing more than 3000 genes) in the span of nearly half a billion years as LUCA survived and evolved in the harsh crater vent environments and gave rise to two distinct domains, bacteria and archaea, documented by microfossil records of the Pilbara and Kaapvaal Cratons (~3.5 Ga). Bacteria and archaea swapped genetic information by horizontal gene transfer, which led to the genetic enrichment and rapid evolution of these two domains [16, 17].

Cumulative fossil evidence from the Nuvvuagittuq Craton, Isua Craton, Pilbara Craton, Kaapvaal Craton, and Singhbhum Craton offer critical information about the origin and ecology of Archean life. Surprisingly, three broad groups of the fossil microbes known from the Archean fossil record—hyperthermophiles, anoxygenic phototrophs, and oxygenic phototrophs—survive today in hydrothermal settings and provide critical information about the habitats of early microbial organisms. For example, some hyperthermophilic communities in Yellowstone National Park can grow into thick mats. Within those mats, hyperthermophile species migrate to the top or bottom depending on temperatures and other conditions. This migratory behavior of early hyperthermophiles in hydrothermal crater vent environments may have given rise to photosynthetic bacteria when they migrated to the water surface to trap sunlight.

Likewise, various groups of anoxygenic photosynthetic bacteria currently thrive in extreme pH, temperature, or salinity conditions. These microbes are uniquely adapted to the conditions of their habitats and are ideal models for defining the physicochemical limits of photosynthesis. For example, in the Mammoth Hot Springs in Yellowstone National Park, purple sulfur bacteria produce laminated microbial mats; these microbes were the precursors of cyanobacteria. They are anoxygenic photosynthesizers, so they harness energy for their metabolism from light.

Finally, there are many examples of cyanobacteria presently living in geothermal habitats, such as the Haughton Crater Lake in Canada and the Lonar Crater Lake in India; these hydrothermal craters display cyanobacterial diversity in both pelagic and benthic habitats [56, 58]. All these examples of living bacteria in geothermal settings help reconstruct the likely habitats of hyperthermophiles, thermophiles, and mesophiles with decreasing thermotolerance in the Archean (Fig. 17.6).

18.4.1 The Habitat of Early Microbial Mats

We reconstruct three distinct microbial regimes from the Archean fossil records, from the bottom to the top. In the earliest stage, microbial mats associated with hyperthermophilic bacteria and archaea appeared in the early Archean around the vent of the hydrothermal crater basin (Fig. 18.2). These hyperthermophiles harnessed chemical energy stored in iron and hydrogen sulfide from the vent environment. The hyperthermophilic nature of ancestral life is supported by an experimental study [59]. A recent phylogenetic analysis has indicated that the earliest enzymes in hyperthermophiles fully adapted to those hot environments [60].

Sunlight became a useful and abundant energy source if there was a pigment to absorb photons. The acquisition of new photosynthetic pigments was a crucial process for the evolution of photosynthesis. In this process, pigment-binding proteins must evolve to create new pigments. When a pigment molecule absorbs photons, they interact with the electronic structure of the bonds holding the molecule together, thereby adding energy to the bonds. Two kinds of pigments evolved in succession: bacteriochlorophyll for anoxygenic photosynthesis and chlorophyll for photosynthetic bacteria [61].

Hyperthermophiles gave rise to two kinds of primitive photosynthetic organisms in succession: anoxygenic photosynthetic bacteria in a thermophilic environment, which then adapted to a mesophilic environment. Nisbet and Fowler [62] suggested that thermotaxis, a behavior in which an organism directs its locomotion up or down a gradient of temperature, led to the evolution of photosynthesis. Microbes moved from the hyperthermophilic condition on the floor of the crater basin to the surface where the primitive, anoxygenic, photosynthetic bacteria (such as purple and green sulfur bacteria) adopted a new frontier in a hydrogen–sulfur habitat. These anoxygenic microbes initially harnessed infrared thermal radiation from the dark hydrothermal vents below the photic zone. Chemosynthetic microbial ancestors used light-sensing molecules. Geothermal light in dark and deep environments paved the way for the evolution of photosynthesis from chemosynthetic microbes. Some of the mobile microbes moved to the surface of the hydrothermal systems, where they harnessed solar radiation, using a new kind of pigment called bacteriochlorophyll during anoxygenic photosynthesis. Green sulfur bacteria could tolerate sulfur, but they are highly oxygen-sensitive. In contrast, purple bacteria develop a tolerance for oxygen and occupy the position below the cyanobacterial mat.

In the final stage of microbial evolution, cyanobacteria emerged on the water surface of the crater lake, adapting to a normal-water temperature in a mesophilic environment. As they tapped the visible solar energy using a more derived pigment, chlorophyll, cyanobacteria invaded the global oceans and produced oxygen. This was a remarkable ecological shift of early microbes from a terrestrial to a shallow marine

environment using abundant visible light as an energy source captured by chlorophyll, when LUCA gave rise to planetary biosphere [63]. The evolution of cyanobacteria, using chlorophyll for metabolism and releasing oxygen as a byproduct, was crucial in creating the modern biosphere. For the first time, oxygen began to accumulate in water and the atmosphere. The emergence of cyanobacteria was the culmination of a long chain of Archean microbe evolution.

We conclude that three ancient microbes adapted thermotolerance in vertically stratified temperature regimes to make use of specific nutrients and energy sources (Fig. 18.1). The cyanobacterial invasion on ocean surfaces resulted in the invasion of cyanobacteria and associated enzymes to adapt to the mesophilic environment [60].

Anoxygenic and oxygenic photosynthetic bacteria created distinctive stromatolite horizons in the Pilbara and Kaapvaal sequences, indicating their microbial activity and specific niches. Phylogenetic models suggest that anoxygenic photosynthesis predates oxygenic photosynthesis [23]. As documented by stromatolites, microbial mats provide a unique ecological niche representative of early life. Cyanobacteria contributed to the oxygenation of the ocean and atmosphere. The availability of oxygen triggered a new kind of oxygen-dependent life, such as respiratory bacteria. The first organisms to 'breathe' oxygen—or at least use it—appeared 3.1 billion years ago, long before it was abundant.

Physiological differentiation and ecological opportunity helped early bacteria and archaea to explore novel environments, from benthic to planktonic, with decreased thermotolerance. These early microbes probably developed chemotaxis in the hydrothermal vent environment to move in response to chemical or photosynthetic stimulus. They must be able to respond to a changing environment, and one way to respond to move from bottom to the surface of the crater lake and finally to the adjacent sea with the development of flagella. In bacteria, a physiological change may be affected by a single-gene acquisition, perhaps mediated by horizontal gene transfer (HGT) and mutation [64]. Genes acquired by

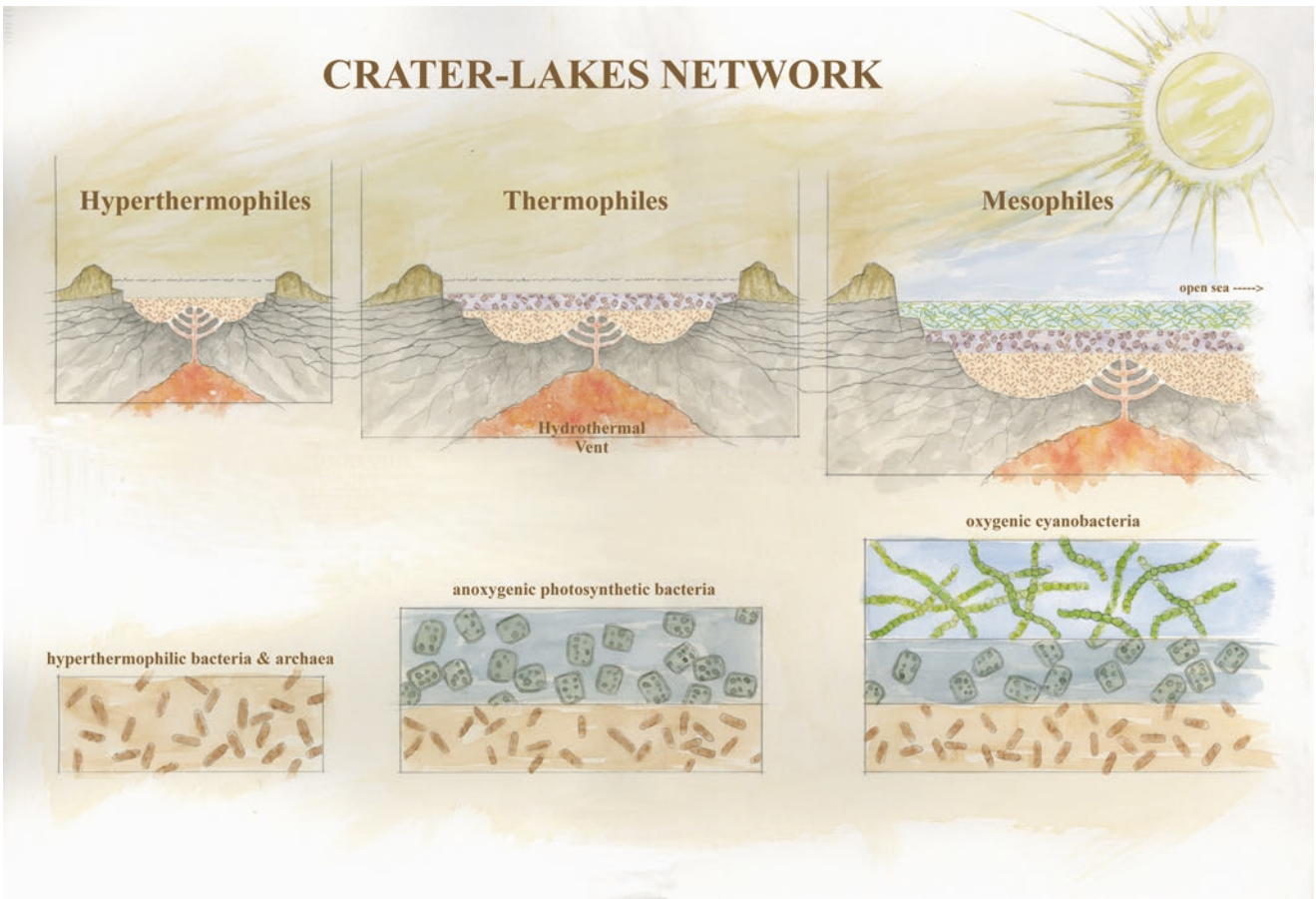


Fig. 18.1 Top, cross sections of three hypothetical, closely spaced hydrothermal crater lakes showing (1) interconnected underground fissure networks and (2) the vertical niches of three microbial communities in the crater lakes. At the bottom of the lakes, close to the vent, hyperthermophilic bacteria and archaea emerge first. Above that, the thermophilic, anoxygenic photosynthetic bacteria evolve. Finally, on the surface, cyanobacteria begin to tap visible sunlight for photosynthesis. As the crater lakes at the margin of the continent merge with the ocean within the top mesophilic zone, cyanobacteria become globally distributed. Bottom, those three microbial zones are enlarged. (Modified from Chatterjee [63])

HGT allowed explore of a new source of abundant energy, i.e., solar energy, and expanded their niches on the global ocean's surface. Ecological stratifications by thermotolerance and different energy sources in the hydrothermal crater lakes allow for genetic isolation and rapid speciation of the Archean microbial community (Fig. 18.1).

18.4.2 Metabolic Evolution of Archean Microbes

Early Archean microbes were essentially autotrophs deriving some or all their organic carbon and nutrients from inorganic sources. They could manufacture their own food (such as sugars and carbohydrates) and tap energy from chemical compounds or sunlight. Two major kinds of autotrophs appeared in the Archaean: chemoautotrophs, such as hyperthermophiles, obtained chemical energy from hydrothermal vent environments, whereas photoautotrophs, such as sulfur bacteria and cyanobacteria, tapped solar energy. For example, autotrophs are rich in organic carbon ^{12}C compared with atmospheric CO_2. As discussed earlier, chemofossils from the Eoarchean deposits provide reliable clues to the presence of autotrophs on young Earth, as revealed from the isotopic composition. In the early Archean biosphere, three kinds of autotrophs existed: hyperthermophilic microbes, anoxygenic photosynthetic bacteria, and oxygenic photosynthetic bacteria. Much of our understanding of the metabolic evolution of Archean microbes is extrapolated from extant life, but the speculation is tentative at best.

18.4.2.1 Chemoautotrophic Hyperthermophiles

The hydrothermal rocks containing the earliest record of life document the presence of anaerobic Earth [62]. Because the photic zone was vulnerable to significant asteroidal impacts during the LHB, life must have originated in a sheltered benthic environment of hydrothermal crater lakes. This was the time interval when the first cellular life and the LUCA originated in harsh environments [66]. Molecular phylogeny suggests that the LUCA was hyperthermophilic [15, 18]. Small hydrothermal crater lakes may be the oldest ecosystems on Earth; this is consistent with the hyperthermophilic origin of life [19, 20, 63]. Morphologically, it would be difficult to distinguish bacteria from archaea from microfossils; archaea no doubt appeared along with bacteria in the early fossil record.

The oldest fossil evidence suggests that hyperthermophilic microbes existed at least 3.8 billion years ago or even earlier [11]. The fossil record from the Pilbara and Kaapvaal Cratons (~3.5 Ga) suggests that life was diversified into two distinct domains, bacteria and archaea, around 3.5 Ga. Archaea were more derived than bacteria [17]. Some bacteria innovated a new kind of metabolism, harnessing solar energy when anoxygenic photosynthetic bacteria and oxygenic photosynthetic bacteria appeared in quick succession (Fig. 18.2). Microbes were exclusively the only forms of life for more than one billion years. There was a progressive increase of genes from LUCA (~500 genes) to cyanobacteria (~3000 genes) [64]. Bacteria and archaea swapped genes by HGT, which led to the rapid evolution of these two domains.

Fossil evidence from the Kaapvaal Craton, the Pilbara Craton, the Nuvvuagittuq Craton, and the Isua Craton provide clues to the origin and ecology of Archean life. Surprisingly, fossil microbes, including hyperthermophilic bacteria and archaea, thermophilic anoxygenic phototrophs, and mesophilic oxygenic phototrophs, are known from the Archean fossil record, and all survive today in hydrothermal settings. They provide critical information about the lifestyle and thermotolerance of early microbial organisms. Some hyperthermophilic colonies in the Yellowstone National Park provide some clues about their migratory behavior; these bacteria produce thick mats. Hyperthermophiles may migrate up or down within those mats, depending on the temperatures of water and air and pH conditions. We can reconstruct this migratory behavior of early hyperthermophiles in the hydrothermal crater vent environment, which may shed light on the rising of the photosynthetic bacteria near the water surface to tap solar energy.

Some anoxygenic photosynthetic bacteria live in extreme temperature, pH, or salinity conditions. These anoxygenic phototrophs are ideal models for defining the physicochemical tolerance of ancient photosynthetic microbes. For example, the Mammoth Hot Springs of the Yellowstone National Park, rich in carbonates and sulfide, exhibit laminated microbial mats; these mats are produced by purple sulfur bacteria. They are anoxygenic photosynthesizers and harness energy for metabolism from infrared. They use hydrogen sulfide in chemosynthetic reactions and produce sulfur. There are many examples today of cyanobacteria living in geothermal habitats: the hydrothermal Haughton crater lake in Canada and the Lonar Crater Lake in India [21, 57, 58]. These bacteria living in a geothermal setting shed light on the ancient habitats of early microbes in the Archean (Fig. 18.1).

Once life evolved, these early microbes dispersed from one hydrothermal crater lake to its neighbor and finally to the sea by an underground network (Fig. 18.1). Although the identification of ancient microfossils by morphology alone appears to be extremely difficult, the molecular phylogeny of microbes and their habitats for early life may provide clues to their identity. Microbial fossils and enclosing sediments from the Archean are consistent with hyperthermophilic adaptation. These microfossils show similar morphology to extant prokaryotes. Similarly, carbonaceous matter from the hydrothermal black cherts (~3.5 Ga) of the Pilbara and Kaapvaal Cratons has provided images of living hyperthermophilic microbes [65]. Only hyperthermophilic bacteria and archaea

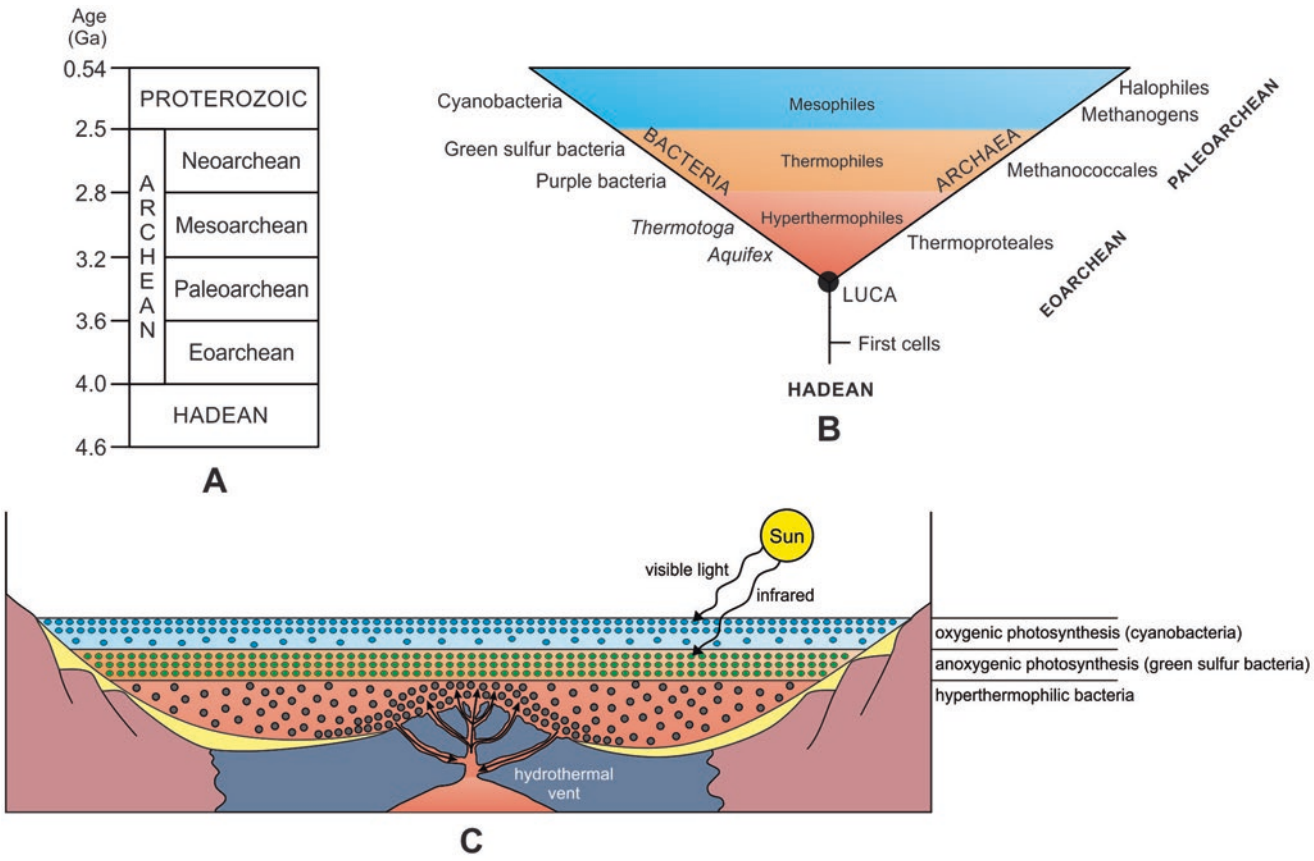


Fig. 18.2 Evolution of the Archean biosphere. (**a**) The geological timescale of the Archean Eon. (**b**), The origin and early evolution of life in an Archean hyperthermophilic benthic crater lake. The two distinct domains of life, bacteria and archaea, are preserved in the fossil record (~3.5 Ga) of the ancient Vaalbara continent. Both domains show the gradual reduction of thermotolerance over time: from hyperthermophilic to thermophilic to mesophilic. In the first hyperthermophilic habitat, both bacteria and archaea appear. In the second thermophilic habitat, bacteria evolve as anoxygenic photosynthesizers. In the final mesophilic habitat, bacteria congregate on the upper surface of the crater lake; they begin to tap solar energy and evolve into oxygenic cyanobacteria. Over time, these cyanobacteria spread globally through the ocean and begin to produce oxygen. On the other hand, hyperthermophilic archaea evolve through two stages: as thermophilic *Methanococcales* and then as mesophilic methanogens and halophiles, with the latter thriving in the hypersaline environment of ponds and lakes. (**c**) A cross section of a hydrothermal crater lake showing the evolution of photosynthesis. The radiation of microbial communities occurs through three stages of an evolving microbial community: first, on the bottom, hyperthermophilic bacteria emerge; next, in the thermophilic middle stage, the anoxygenic photosynthetic green sulfur bacteria appear; and, in the final stage, at the upper mesophilic level, oxygenic photosynthesizing cyanobacteria form and begin the production of oxygen. (Modified from Chatterjee [63])

could survive in the sediments associated with Archean pillow lavas [45]. The early evolution of archaea is supported by the discovery of ^{13}C-depleted methane in hydrothermal fluid inclusions in cherts from the Pilbara Craton. Ueno et al. [38] reported evidence for the existence of an ancient population of methane-producing archaea from the Dresser Formation (~3.5 Ga). The fossil record suggests that bacteria and archaea must have split at least 3.5 Ga or earlier.

It appears that the Archean fossils from the Nuvvuagittuq, Isua, Pilbara and Kaapvaal Cratons provide crucial evidence for three phases of environmental temperatures and niche adaptation over 300 million years: in the first period, hyperthermophilic bacteria and archaea adapted high thermotolerance around the vent of the crater basin, nourished by chemicals from spewing vents. In the second period, the thermophilic anoxygenic photosynthetic bacteria lived near the water surface to tap infrared rays; they used hydrogen sulfide to produce glucose and released sulfur as a byproduct. In the third phase, cyanobacteria adapted to the mesophilic temperature regime on the water surface to harness visible light; they used water to build glucose and produced oxygen as a byproduct. The fossil record suggests that chemosynthesis preceded phototrophy (Fig. 18.1).

18.4.2.2 Photoautotrophs: I. Anoxygenic Photosynthetic Bacteria

The rise of photosynthesis is one of the crucial episodes in the development of Archean microbial communities. Nisbet and Fowler [62] suggested that photosynthesis began as an adaptation of thermotaxis. The origin of photosynthesis is little

understood, but Archean fossils provide some clues. Early hyperthermophiles lived in hydrothermal vents for energy and nutrition. Phototrophy is a process by which organisms trap light energy (photons) and store it as chemical energy in the form of adenosine triphosphate (ATP). Most likely, photosynthetic organisms were present in the younger sequences of the Kaapvaal and Pilbara Cratons (~3.4–3.2 Ga) in the form of stromatolites and microfossils, which are inferred from a morphological or geological context [40, 53, 65, 66].

Photoautotrophic bacteria exhibit two kinds of biochemistries and metabolisms [61, 62]. The primitive anoxygenic photosynthetic bacteria absorbed near-infrared rather than visible sunlight and produced sulfur or sulfate compounds rather than oxygen. They use molecules such as H_2S instead of H_2O for photosynthesis, and water is not employed as an electron donor. H_2S is supplied to the vent environment by geothermal activity. Their pigments (possibly bacteriochlorophylls) were predecessors to chlorophyll. Anoxygenic phototrophs and their metabolism produce sulfur as a byproduct and build glucose, the universal cellular fuel. They use infrared instead of visible light for metabolism.

$$\left(\text{Anoxygenic photosynthesis}\right): H_2S + CO_2 \rightarrow CH_2O + S$$

The origin of photosynthesis is poorly understood. Xiong et al. [64] obtained new sequence information from the green sulfur bacterium *Chlorobium* and the green non-sulfur bacterium *Chloroflexus* and conclusively demonstrated for the first time that the major lineages of the pigment (bacteriochlorophyll) involved in anoxygenic photosynthesis arose before the development of oxygenic photosynthesis (chlorophyll) [66, 67].

18.4.2.3 Photoautotrophs: II. Oxygenic Photosynthetic Bacteria

By the late Archean, a more complicated form of photoautotrophy, i.e., oxygenic photosynthesis, evolved, as manifested by cyanobacteria associated with their stromatolites. In cyanobacterial synthesis, the pigment is chlorophyll, hydrogen is always provided by water, and glucose and oxygen are given off as byproducts.

$$\left(\text{Oxygenic photosynthesis}\right): H_2O + CO_2 \rightarrow CH_2O + O_2$$

Solar radiation was overwhelmingly the new energy source for cyanobacteria to access different chemistries rather than the energy provided by other sources, such as hydrothermal vents, because of the unique characteristics of photochemistry that differentiate it from conventional hyperthermal chemistry. The ecological transition of the Archean life from the benthic thermophilic life to anoxygenic photosynthetic bacteria in the crater vent environment and finally to planktonic mesophilic cyanobacteria in the shallow marine environment was rather rapid, as documented in the Pilbara and Kaapvaal Cratons (Fig. 18.2). However, we are left with the possibility that the photic (sunlit) zone could have been destroyed intermittently throughout the Archean Eon, as revealed by the spherule ejecta layers between 3.5 and 3.2 Ga [34]. Perhaps, after the cessation of significant bolide impacts after 3.2 Ga, obligate photosynthetic cyanobacteria could continuously evolve. This is the time when the paleontological record indicates highly evolved photosynthetic systems, transforming the atmosphere and permitting the evolution of eukaryotes. The Archean fossils from South Africa and Western Australia shed new light on photosynthesis's origin and early evolution from their hyperthermophilic bacterial ancestors.

One of the most critical developments in Earth's history is the change from the anaerobic environment of early Earth to the aerobic and highly oxidizing environment that we have today, with 21% atmospheric oxygen. Although the 'Great Oxidation Event' that led to the oxygen bloom occurred on our planet in the early Proterozoic (~2.3 Ga) with the proliferation of cyanobacteria, chlorophyll-based photosynthesis originated early in the bacterial domain in the Archean (~3.2 Ga). This led to the onset of global oxygenation in the atmosphere–biosphere system in the early history of Earth. The accumulation of oxygen in our atmosphere was an exceedingly protracted process. It is still an enigma how early cyanobacteria developed the ability to oxidize water and produce oxygen as a byproduct. In photosynthesis, light energy causes chlorophyll to enter a high-energy excited state that releases an electron. The electron travels through a series of reactions to carbon dioxide, which is reduced to sugar such as glucose. In a second reaction, the chlorophyll recovers an electron from water with the result that oxygen is released. Photosynthesis produces all oxygen in the atmosphere. It generates a distinctive chemical signature in the form of a carbon isotope composed of the organic material produced. During photosynthesis, ^{12}C (carbon-12 in CO_2) is selectively absorbed by the photosynthetic machinery, compared with ^{13}C (carbon-13), that is, it reacts faster with the enzymatic system and is therefore selectively incorporated into the organic compounds produced. The carbon-12 isotopic ratio has been used to trace the early evidence of life in Greenland (~3.8 Ga) [26].

Phylogenetic studies of 3983 gene families across the three domains of life plotted along a geological timeline suggest the rapid evolutionary innovation of microbial life during a brief period in the Archean Eon around 3 Ga, which coincides with a rapid diversification of bacterial lineages and genetic expansion; the microbial gene expansion was mediated by gene duplication and horizontal gene transfer. Genes arising after this expansion show the increasing use of molecular oxygen and redox-sensitive transition metals such as iron and nickel [68].

18.5 Radiation of Archean Microbial Communities

Microbes played critical roles in the biogeochemical cycles that drive our planet. Life may have initially started on the crater floor of deep and dark hydrothermal vent environments in an anoxic environment. Still, life would ultimately spread onto vast sunlight-exposed ocean surfaces. Early microbial evolution refers to that phase of biological history that began with the emergence of life forms such as hyperthermophiles, followed by purple and green sulfur bacteria in the anaerobic environment and the explosive diversification of aerobic photosynthesizers such as cyanobacteria. Initially, hyperthermophiles and anoxygenic photosynthesizers clustering around hydrothermal vents may have been great contributors to microbial metabolisms in the early Archean. Eventually, oxygenic photosynthesis would evolve and eclipse its anoxygenic cousins, thus becoming the driver of the global carbon cycle [69] (Fig. 18.2).

The development of oxygenic photosynthesis by cyanobacteria around 3.2 billion years ago, or even earlier, transformed Earth forever. This grand biogeochemical shift set in motion the evolution of subsequent microbial metabolisms and lifestyles. Oxygenic photosynthesis triggered the abundance of life. Once life emerged on Earth, it proliferated across the global ocean by mutation through Darwinian evolution. The ability of oxygenic photosynthetic bacteria to capture hydrogen by splitting water was an extraordinary innovation in the microbial community. The virtually unlimited supply of hydrogen in water freed life from its sole dependence on the abiotic chemical sources found in hydrothermal vents. Communities sustained by oxygenic photosynthesis could thrive wherever supplies of sunlight, water, and nutrients were sufficient. Three billion years later, photosynthesis continues to play a vital role in regulating atmospheric levels of oxygen and carbon dioxide. Photosynthetic microbial communities have left a relatively robust fossil record because of their extremely high productivity on stable continental shelves; this contributed to sediments with excellent potential for long-term preservation. Although cyanobacteria emerged during the mid-Archean or even earlier, their microfossil record is robust throughout the Proterozoic.

The oxygen, as the result of the photosynthetic activity by cyanobacteria, was so pervasive and so productive that it eventually reached levels sufficient to drive the emergence of respiratory bacteria, eukaryotes, and complex multicellular life in the Proterozoic. The endosymbiotic merging of cyanobacteria with respiratory bacteria drove a complex eukaryote biosphere [69]. The origin of eukaryote cells is considered a milestone in the evolution of life because they comprise all multicellular organisms—plants, fungi, animals, and us. Without eukaryotes, we would not be here to discuss the question of the origin of life. The fossil record does not provide many clues about the time of their birth. The oldest fossil record of eukaryotes goes back to 2.7 billion ago, with the discovery of eukaryotic biomarkers in ancient oil. The oldest eukaryotic body fossil is *Grypania spiralis* from the 2.1-billion-year-old Negaunee Iron Formation of Michigan. Molecular evidence suggests that three domains of life—bacteria, archaea, and eukaryotes—appeared in the Archean Eon [68]. For eukaryotes to thrive, oxygen had to be present in the atmosphere in higher amounts than what existed on early Earth. A great diversity of microbial and aerobic habitats must have lived in the late Archean. Eukaryotes are more closely allied to archaea than bacteria in nuclear DNA and genetic machinery [16, 17].

One of the earliest records of Archean microbial ecosystems comes from stromatolites (~3.5–3.2 Ga), which display structures like modern microbial mats. The earliest microbial mats consisting of bacteria and archaea may have formed as biofilms in hydrothermal vents. As photosynthetic bacteria left hydrothermal vents and invaded the shallow ocean, the microbial mats developed profusely in this environment, allowing widespread colonization of the globe and creating different aerobic habitats. A key feature of modern-day microbial mats and stromatolites is that sunlight controls the orientation and stratification. Microbial mats are a unique ecological niche representative of early life on Earth. Rapid nutrient cycling across micro-gradients coupled with putative niche differentiation within mat layers enables diverse metabolic processes to occur in spatial proximity.

On a broad scale, cyanobacteria's origin of oxygenic photosynthesis led to the rise of oxygen on Earth during the early Proterozoic (~2.4–2.3 Ga), which played a significant role in the oxygenation of the ocean and atmosphere and paved the way for oxygen-breathing microbes such as respiratory bacteria. The 'Great Oxidation Event' is when the atmosphere first became oxygenated [70]. However, oxygen levels only reached somewhere between 0.2% and 2% by volume, not today's 21% [18]. Respiration is the opposite of photosynthesis. Whereas photosynthesis harnesses energy from the Sun, respiration releases energy. In photosynthesis, cyanobacteria combine carbon dioxide and water to form glucose and oxygen. In respiration, aerobic bacteria consume sugar and oxygen to gain energy. A single biochemical trick that evolved around 3.1 Ga to produce oxygen is widely being used for respiration by all animals, including humans.

The history of life is mostly microbial. Microbes once ruled the whole world. Earth has been a bacterial world for at least the last 3.5 billion years. The diversity of life today is a result of the dynamic interplay between genetic opportunities, metabolic capabilities, and environmental changes. Earth's habitable environments have been dominated by

microbes and subjected to their metabolism and evolution for most of their existence. Prokaryotes gave life its initial foothold. Photosynthetic bacteria make our planet habitable for us, inhaling carbon dioxide and exhaling oxygen, day in and day out, for billions of years, building enough oxygen in the water and atmosphere to support more abundant and complex life [70].

18.6 Conclusions

Eoarchean hydrothermal vent environments provided stable, anaerobic, and hot habitats in which hyperthermophilic life could have emerged and become established. These hydrothermal fields were geochemically reactive in the far-from-equilibrium setting and contained an assortment of cosmic and terrestrial organic compounds. Hydrothermal vents were powered by volcanic, chemical, and solar energies and created convection cells that drove the prebiotic synthesis to form more complex organic molecules, initiating molecular symbiosis and providing physically and chemically diverse and secured habitats for microbial life. The widespread occurrence of hydrothermal crater lakes on the Eoarchean crust has led to the suggestion that such environments may have constituted the crucible of life on Earth. It was on this chaotic and violent Earth, before the onset of plate tectonics, around four billion years ago, that the first life probably emerged in hydrothermal crater lakes. The Archean atmosphere was anoxic and reducing, mainly consisting of CO_2 and N_2.

The first life probably arose on Earth about four billion years ago at the time of the Hadean–Archean transition. The molecular concept of the origin of life suggests that life originated much earlier (~4.4 Ga) in the early Hadean, soon after the Moon-forming impact. Although no Hadean rocks are preserved, detrital zircons from Australia and India indicate the presence of water in the Hadean. The first billion years of biological and environmental evolution featured bacteria and archaea in the hydrothermal crater lakes. The earliest fossil record (≥3.8 Ga) of biotic activity is preserved in the Archean hydrothermal sedimentary rocks of Canada, Greenland, Australia, South Africa, and India in the form of carbonaceous remains of microbial cells, cellular microfossils, and stromatolites. In the younger sequences of these Archean cratons, photosynthetic bacteria, anoxygenic green sulfur bacteria, and oxygenic cyanobacteria appeared quickly from their hyperthermophilic ancestors. The universal tree of life shows that both kinds of hyperthermophilic microbes evolved early, but archaea were more derived than bacteria. Hyperthermophiles gave rise to anoxygenic photosynthetic bacteria, which then evolved into oxygenic photosynthetic bacteria. Oxygenic photosynthesis allowed life to escape the hydrothermal setting and invade an utterly new frontier—broad continental shelves. Cyanobacteria contributed to the geological processes by producing vast amounts of carbonate sediments and stromatolitic structures in the shallow seas; they formed oxygen as the byproduct, transforming the ocean and the atmosphere ~3.2 billion years ago and triggering an explosive evolution. Symbiosis of early microbes and viruses probably led to the origin of eukaryotes during the early Proterozoic era.

References

1. Betts HC, Puttick MN, Clark JW, Donoghue PCJ, Pisani D (2018) Integrated genomic and fossil evidence illuminates life's early evolution and Eukaryote origin. Nature Ecol Evol 2:1556–1562
2. Harrison TM, Schmitt AK, McCulloch MT, Lovera OM (2008) Early (>4.5 Ga) formation of terrestrial crust: Lu-Hf, $\partial^{18}O$, and Ti thermometry results for Hadean zircons. Earth Planet Sci Lett 268:476–486
3. Miller SR, Mueller PA, Meert JG, Kamenov GD, Pivarunas AF, Sinha AK, Pandit MK (2018) Detrital zircons reveal evidence of Hadean crust in the Singhbhum Craton, India. J Geol 126:541–552
4. Mojzsis SJ, Harrison TM, Pidgeon TT (2001) Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. Nature 409:178–18115
5. Dodd MS, Papineau D, Grenne T, Slack JF, Rittner M, Prajno F, O'Neil JO, Little CTS (2017) Evidence for early life in Earth's oldest hydrothermal vent precipitates. Nature 543:60–65
6. Papineau D, She Z, Dodd MS, Iacoviello F, Slack JF, Hauri E, Shearing P, Littlr TS (2022) Metabolically diverse primordial microbial communities in Earth's oldest seafloor-hydrothermal jasper. Sci Adv 8(15). https://doi.org/10.1126/sciadv.abm2296
7. Garcia-Ruiz JM, Melero-Garcia E, Hyde ST (2009) Morphogenesis of self-assembled nanocrystalline materials from barium carbonate and silica. Science 323:362–365
8. Sugitani K (2018) Fossils of ancient microrganisms. In: Kolb VM (ed) Handbook of astrobiology. CRC Press/Taylor & Francis, Boca Raton, pp 567–596
9. Catling DC, Zahnle KJ (2020) The Archean atmosphere. Sci Adv 6(9). https://doi.org/10.1126/sciadv.aax1420
10. Tang M, Chen K, Rudnick RL (2016) Archean upper crust transition from mafic to felsic marks the onset of plate tectonics. Science 351:373–375
11. Bell EA, Boehnke P, Harrison TM, Mao WL (2015) Potentially biogenic carbon preserved in a 4.1-billion-year-old zircon. Proc Natl Acad Sci USA 112:14518–14521
12. Schopf JW (1993) Microfossils of the Early Archean Apex Chert: new evidence for the antiquity of Life. Science 260:640–646
13. Brasier MD, Green OR, Jephcoat AP, Kleppe AK, Van Kranendonk MJ, Lindsay JF, Steele A, Grassineau NV (2002) Questioning the evidence of Earth's oldest fossils. Nature 416:76–81
14. Van Zullen MA, Lepland A, Arrhenius G (2002) Reassessing the evidence for the earliest traces of life. Nature 418:627–630
15. Pace NR (1997) A molecular view of microbial diversity and the biosphere. Science 276:734–740
16. Woese CR (1994) The universal ancestor. Proc Natl Acad Sci USA 99:6864–6859
17. Woese CR (2002) On the evolution of cells. Proc Natl Acad Sci USA 99:8742–8747
18. Gaucher EA, Govindan S, Ganesh OK (2008) Paleotemperature trend for Precambrian life inferred from resurrected proteins. Nature 451:704–707
19. Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. Nature Rev Microbiol 1:127–136

20. Osiniski GR, Tornabene LL, Banerjee NR, Cockell CS, Flemming R, Izawa MRM, McCutcheon J, Parnell J, Preston LJ, Pickersgill AE, Pontefract A, Sapers HM, Southam G (2013) Impact-generated hydrothermal systems on Earth and Mars. Icarus 224:347–363
21. Parnell J, Lee P, Cockell CS, Osinski GR (2004) Microbial colonization in impact-generated hydrothermal sulphate deposits, Haughton impact structure, and implications for sulfate on Mars. Int J Astrobiol 3:247–225
22. Reysenbach AL, Cady SL (2001) Microbiology of ancient and modern hydrothermal systems. Trend Microbiol 9:79–86
23. Nisbet EG, Sleep NH (2001) The habitat and nature of early life. Nature 409:1083–1091
24. Fedo CM, Whitehouse MJ (2002) Metasomatic origin of quartz-pyroxene rock, Akilia, Greenland, and implication for Earth's life. Science 296:1448–1452
25. Abramov O, Mojzsis SJ (2009) Microbial habitability of the Hadean Earth during the late heavy bombardment. Nature 449:419–422
26. Mojzsis SJ, Arrhenius G, McKeegan KD, Harrison TM, Nutma AP, Friend CRL (1996) Evidence for life on Earth before 3,800 million years ago. Nature 384:55–59
27. Grassineau NV, Abell P, Appel PWU, Lowry D, Nisbet EG (2006) Early life signatures in sulfur and carbon isotopes from Isua, Barberton, Wabigon (Sleep Rock) and Balingwe Greenstone belts (3.8 to 2.7 Ga). Geol Soc Am Mem 198:33–52
28. Rosing MT (1999) ^{13}C-depleted carbon microparticles in >3700-Ma sea-floor sedimentary rocks from West Greenland. Science 283:674–676
29. Cates NL, Mojzsis SJ (2007) Pre-3750 Ma supracrustal rocks from the Nuvvuagittuq supracrustal belt, northern Quebec. Earth Planet Sci Lett 255:9–21
30. Nutman AP, Bennett VC, Friend CR, Van Kranendrok MJ, Chivas AR (2016) Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. Nature 537:535–538
31. Hedges SB (2002) The origin and evolution of model organisms. Nature Rev Genet 3:838–849
32. De Koch MO, Evans DAD, Beukes NJ (2009) Validating the existence of Vaalbara in the Neoarchean. Precamb Res 174:145–154
33. Zegers TE, De Wit MJ, Dann J, White SH (1998) Vaalbara, Earth's oldest assembled continent? a combined structural, geochronological, and palaeomagnetic test. Terra Nova 10:250–259
34. Byerly GR, Lowe DL, Wooden JL, Xie X (2002) An archean impact layer from the Pilbara and Kaapvaal Cratons. Science 297:125–1327
35. Furnes H, Banerjee NR, Muehlenbachs K, Staudigel H, de Wit M (2004) Early life recorded in Archean pillow lavas. Science 304:578–581
36. Westall F, de Wit MJ, Dann J, van der Gaast S, Ronde CEJ, Gerneke D (2001) Early Archean fossil bacteria and biofilms in hydrothermally influenced sediments from the Barberton Greenstone belt, South Africa. Precambr Res 106:93–116
37. Wacey D, Kilburn MS, Saunders M, Cliff J, Brasier MD (2011) Microfossils of sulfur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. Nature Geosc 4:698–702
38. Ueno Y, Yamada K, Yoshida N, Maruyama S, Isozakai Y (2006) Evidence from fluid inclusions for microbial methanogenesis in the early Archean era. Nature 440:516–519
39. Djokic T, Van Kranendonk MJ, Campbell KA, Walter MR, Ward CR (2017) Earliest signs of life on land preserved in ca, 3.5 Ga hot spring deposits. Nature Commun. https://doi.org/10.1038/ncomms15263
40. Noffke N, Daniel C, Waxey D, Hazen RM (2013) Microbially induced sedimentary structures recording an ancient ecosystem in the ca. 3.48 billion-year-old Dresser Formation, Pilbara, western Australia. Astrobiology 13:1103–1124
41. Schopf JW, Tripathi A, Kudryavatev AB, Czaja AD, Tripathi AB (2007) Evidence of Archean life: stromatolites and microfossils. Precambrian Res 158:141–155
42. Pinti DL, Mineau RM, Clement V (2009) Hydrothermal alteration and microfossil artifacts of the 3,465-million-year-old Apex Chert. Nat Geosci 2:640–643
43. Rasmussen B (2000) Filamentous microfossils in a 3,235-million-year-old volcanogenic massive sulfide deposit. Nature 405:676–679
44. Farmer JD (2000) Hydrothermal systems: doorways to early biosphere evolution. GSA Today 10:1–9
45. Kring DA (2000) Impact events and their effect on the origin, evolution, and distribution of life. GSA Today 10(8):1–7
46. Sapers HM, Osinski GR, Banerjee NR, Preston LJ (2015) Enigmatic tubular features in impact glass. Geology 43:635–638
47. Banerjee NR, Furnes H, Muehlenbachs K, Staudigel H, de Wit M (2006) Preservation of ~3.4–3.5 Ga microbial biomarkers in pillow lavas and hyaloclastites from the Barberton Greenstone Belt, South Arica. Earth Planet Sci Lett 241:707–722
48. Glikson M, Duck LJ, Golding S, Hoffman A, Bolhar R, Webb R, Baiano J, Sly L (2008) Microbial remains in some earliest Earth rocks: comparison with a potential modern analogue. Precambrian Res 164:187–200
49. Walsh AM (1992) Microfossils and possible microfossils from the Early Archean Onverwacht Group, Barberton Mountain Land, South Africa. Precambrian Res 54:271–292
50. Walsh AM, Lowe DR (1985) Filamentous microfossils from the 3,5000-Myr-old Onverwacht Group, Barberton Mountain Land, South Africa. Nature 314:530–532
51. Tice MM, Lowe DR (2004) Photosynthetic microbial mats in the 3,146-Myr-old ocean. Nature 430:549–552
52. Knoll AH, Barghoorn ES (1977) Archean microfossils from the early Precambrian of South Africa. Science 156:508–512
53. Javaux EJ, Marshall CP, Bekker A (2010) Organic-walled microfossils in 3.23-billion-year-old shallow-marine siliclastic deposits. Nature 463:934–938
54. Basu AR, Bandopadhyay PK, Chakrabarti R, Zou H (2008) Large 3.4 Ga Algoma-type BIF in the eastern Indian Craton. Geochem Cosmochem Acta 72:A59
55. Mazumdar R, Chaudhury T (2021) Paleoarchean terrestrial to shallow marine sedimentation on Singhbhum Craton, eastern India (the Western Iron Ore Group). Precambrian Res 354. https://doi.org/10.1016/j.precamres.2020.106071
56. Maithy PK, Kumar S, Babu R (2000) Biological remains and organo-sedimentary structures from Iron Ore Supergroup (Archaean) Barbil area, Singhbhum, Orissa. Geol Surv India Spec Publ 57:98–106
57. Hagert JJ, Newsom HE (2003) Hydrothermal alteration at the Lonar Lake impact structure, India: implications for impact cratering on Mars. Meteor Planet Sci 38:365–381
58. Paul D, Kumbhare SV, Mhatre SS, Chowdhury SP, Shett SA, Marathe NP, Bhute S, Shouche YS (2016) Exploration of microbial diversity and community structure of Lonar lake: the only hypersaline meteorite crater-lake within basalt rock. Front Microbiol. https://doi.org/10.3389/fmicb.2015.0155
59. Akanuma S, Nakajima Y, Yokobori S, Kimura M, Nemoto N, Mase T, Miyazono K, Tanokura M, Yamagishi A (2013) Experimental evidence for the thermophilicity of ancestral Life. Proc Natl Acad Sci 110:11067–11072
60. Nguyen V, Wilson C, Hoemberger M, Stiller JB, Agavonof RV, Kutter S, English J, Theobold DL, Kern D (2017) Evolutionary drivers of thermoadaptation in enzyme catalysis. Science 355:289–294
61. Schopf JW (1999) Cradle of life: the Discovery of Earth's Earliest Fossils. Princeton University Press, Princeton
62. Nisbet EG, Fowler CMR (1996) Some liked it hot. Nature 382:404–405
63. Chatterjee S (2018) The hydrothermal impact crater lakes: the crucibles of life's origin. In: Kolb VM (ed) Handbook of astrobiology. CRC Press/Taylor & Francis, Boca Raton, pp 265–295

64. Doolittle WF (2000) Uprooting the tree of life. Sci Am 282(2):90–95
65. Westall F, Campbell KA, Breheret JG, Foucher F, Gautret P, Hubert A, Sorfieul S, Grassineau N, Guido DM (2015) Archean (3.33 Ga) microbe-sediment systems were diverse and flourished in a hydrothermal context. Geology 43:615–618
66. Blankenship RE (2010) Early evolution of photosynthesis. Plant Physiol 154:434–438
67. Xiong J, Fischer WM, Inoue K, Nakahara M, Bauer CE (2000) Molecular evidence for the early evolution of photosynthesis. Science 289:1724–1730
68. David LA, Alm EJ (2011) Rapid evolutionary innovation during an Archean genetic expansion. Nature 469:93–96
69. Margulis L (1981) Symbiosis in cell division. Freeman, San Francisco, CA
70. Arndt NT, Nisbet EG (2012) Processes on the young earth and habitats of early life. Annu Rev Earth Planet Sci 40:521–549

Life Beyond Earth

19

The sky, adorned with stars and Sun
The Universe, teeming with life
Within this glorious vast expanse
Here I am, with a place to be, a role to play
I am amazed and wondering,
I see my song find itself here.

—Rabindranath Tagore, 1924

19.1 Are We Alone in the Universe?

As the stars move across the sky each night, humans have looked up and wondered about their place in the universe. Are we alone? Is Earth the only life-sustaining planet? Are there others like us out there? Does life, be it like our own or different, exist elsewhere in the solar system? Our galaxy? Our universe? Our ongoing discovery of extraordinary new worlds, including Earth-like exoplanets, suggests that the universe may be teeming with life. However, the big question is: will we ever find extraterrestrial intelligence?

The quest for life beyond our planet is a dream as old as humankind itself. The ancient Greeks were the first Western thinkers to recognize the possibility of an infinite universe or a multiverse housing an endless number of civilizations. In 600 BC, Thales espoused the idea of a plurality of living worlds. Subsequently, the concept of alien worlds was opposed by Plato and Aristotle. They suggested that Earth was uniquely inhabited and lay at the center of the universe. This theory prevailed for more than a thousand years.

Explaining Epicurean philosophy to a Roman audience more than 2000 years ago, the poet and philosopher Lucretius wrote in his *De Rerum Natura* (*On the Nature of Things*) that ours cannot be the only inhabited world: 'Confess you must that other worlds exist in other regions of the sky, and different tribes of men, and kinds of wild beasts.' Much later, in the sixteenth century, Copernicus showed that Earth is not the center of the universe but just one of the planets in the solar system and that other worlds are possible just like ours. We may not be the only planet with life, but perhaps just one of many. By the early seventeenth century, German astronomer Johannes Kepler speculated in his book *Somnium* (*The Dream*) that there might be life on other worlds and spread throughout the universe. The Italian astronomer Giordano Bruno argued that if the Sun has a planetary family, so too might other stars, with their planets equally capable of supporting life. For his heretical view, in 1600, the papal authority convicted and burned Bruno alive at the stake in Rome's Campo de' Fiori. Bruno's hypothesis about the plurality of worlds turned out to be true with the discovery of exoplanets by NASA.

However, it was not until the close of the 1950s that anyone proposed a credible method to look for these distant hypothetical neighbors. The space age had dawned, and science could now provide new tools to explore what lay beyond Earth. Since then, the question of whether alien life exists has fired the imagination, fed speculation, and fueled debate among scientists, science fiction writers, and moviemakers alike. As the spacecraft 'Voyager 1' left our planetary neighborhood in 1990, NASA engineers turned it around for one last look at the home planet and captured a portrait of our world. Earth appeared 6.4 billion kilometers away, at the solar system's outer edge, as a tiny point of light. Are we alone—a 'pale blue dot' in a sunbeam bursting with life amidst a vast, uninhabited universe—or is there more life beyond our shores?

19.2 Astrobiology

Much of the universe is inhospitable to life, and only rare places offer potential sanctuaries for its existence. Life is a planetary phenomenon. Our planet Earth is unique because of its habitable or Goldilocks location in the solar system, just right for water to remain a liquid, and its abundant biodiversity. We began this inquiry by asking why Earth is the

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_19

only planet in our solar system with abundant life. Earth is habitable for a number of factors: it is the right distance from the Sun; it has the right size for optimum atmospheric pressure, allowing water to be a liquid on its solid surface; and it is big enough to have acquired a vast store of volatile gases that can be vigorously degassed and recycled by its plate tectonic system. The critical ingredients for life as we know it on Earth are liquid water and organic compounds. However, we now know that these stocks are also abundant in interstellar space and meteorites.

The quest for extraterrestrial life is a conundrum that has fueled incredible scientific feats and profound popular culture. Astrobiology has emerged as a new branch of science intent on studying the origin, evolution, and future of life in the universe, out in the stars, and here on Earth. Measured from edge to edge, the universe as we know it stretches some 93 billion light-years across. This unfathomable expanse contains two trillion galaxies, each shining with millions of stars and dotted with more planets than we can imagine. This vast universe contains something like 10^{21} stars. In the Milky Way galaxy, there are between 250 and 500 billion stars.

NASA's research provides guidelines for our future search for alien life—what we know, what we do not know, and what we might hope to learn about potential life on distant planets. Since the 1950s, scientists have argued that the 'habitable zones' around stars are the most probable places to find life. Beyond the solar system, the Kepler space telescope has discovered more than 5000 extrasolar planets, or 'exoplanets,' orbiting the habitable zones of Sun-like stars and red dwarfs in the Milky Way alone. Most known exoplanets orbit stars roughly like the Sun. Some of the exoplanets are probably habitable and may hopefully harbor life. The question is no longer is there any life beyond Earth, but, rather, how do we find it?

There is a strong possibility that life has also emerged beyond Earth whenever the necessary physical and chemical conditions were met. De Duve [1] claims that the constraints of physics and chemistry are so strong that life at any other place beyond Earth must be the same as it is here. He believes that it is not accidental that a limited number of organic compounds formed under physical and chemical conditions on early Earth and that these organic compounds reacted among themselves and created life. Although the details regarding amino acids and nucleotides must have differed, it is believed that life elsewhere will still be based on organic chemistry (carbon and water) and a genetic information system. NASA's Astrobiology Program has established the expectation of microbial life in many other habitable places within and beyond our solar system. Most astrobiologists believe that microbial life is widespread in the solar system, the galaxy, and the universe and may have the same biological basis as Earth. However, significant degrees of contingency are associated with the nature of more complex organisms and modes of intelligence.

One of the most critical developments in this regard is discovering that microbial life is abundant and diverse in extreme environments on Earth. To begin with, Earth's oceans are unique in the solar system. No other world orbiting our Sun has a surface covered with liquid water. By contrast, icy oceans appear to be relatively common in the outer solar system. Nevertheless, there are at least 10 celestial bodies in our solar system that might harbor life: Mars, three large icy moons of Jupiter, namely, Europa, Ganymede, and Callisto, Saturn's icy moon Enceladus and its largest moon Titan, Neptune's moon Triton, the Kuiper Belt object Pluto, and two asteroids, namely, Ceres and Bennu. Extremophiles' wide-ranging adaptations to high-pressure, dark, toxic, and/or salty niches as well as to alkaline or acidic environments alike may offer the most important clues to the possibility of extraterrestrial life in the inhospitable habitats of Mars, Jupiter's moon Europa, and Saturn's moons Titan and Enceladus. There is not yet any direct evidence of life on other planets in our solar system or galaxy. However, recent NASA explorations have prompted optimism that habitable worlds exist throughout the Milky Way. However, if so, how do we detect it? How do we find those specific habitats on other planets where life may exist, and how do we get to them?

With the discovery of extremophiles, we have revised the concept of habitable zones to include other categories, which now accommodate Europa, Titan, and Enceladus, which are far from the Sun. Life is exceptionally resilient once it takes hold, but it requires rich chemistry, large energy resources, and stability. As we have discussed earlier, if meteorites delivered the building blocks of life to Earth, then they could surely deliver them to other planets and their circling satellites. Therefore, life potentially began in other habitable niches or extreme aquatic environments throughout the solar system. Although Mercury, Venus, and the Moon are undoubtedly bereft of life now, each has features that imply that earlier in their history, conditions may have been sufficiently favorable to support life.

The success of recent NASA space probes has provided us with a wealth of new data about the planets in our solar system. However, although we can detect habitable planets outside our solar system with space telescopes, we do not yet have the technology to travel to them. In the meantime, NASA's orbiting Hubble Space Telescope has detected atmospheres and identified their compositions on many planets outside our solar system, so we can now search them for chemical markers of life or 'biosignatures.' Similarly, NASA's Kepler space telescope has discovered several Earth-like planets in the habitable zones around other stars. Still, it cannot yet tell us whether life exists there now or has existed there in the past.

Several different biosignatures can be used to detect the conditions for life, past or present, on Earth and elsewhere. This list includes microbial metabolic gases, such as methane and oxygen, spectral techniques for analyzing atmospheres, biomolecules, and their degradation products, homochiral organic molecules, isotopic fractionation (including C, S, N, and other elements), biominerals formed by microbial metabolism (such as sulfur, carbonate, and phosphate), the morphological features produced by microorganisms, bacterial fossils, and many more. Thus, astrobiologists must decide which biosignatures to target when searching for signs of life on faraway planets.

Thousands of exoplanets, or planets orbiting stars like our Sun, have been discovered in the past decade. According to NASA, the gold standard in biosignature gases to target alien life on faraway planets is oxygen. Oxygen is a trademark of life on Earth, and it is likely to become widespread in the biosphere. Not only is oxygen produced in abundance by Earth's photosynthesizers—cyanobacteria, algae, and plants—and thus, possibly, other planets'—but it could not be produced at detectable levels by geology or photochemistry alone, making it a reliable sign of alien life.

19.3 Searching for Life in the Solar System

The discovery of Earth's early microbial fossils has clarified the beginnings of biological evolution in watery environments. NASA's strategy in looking for life in the solar system is to 'follow the water.' Our two nearest neighbors on either sides, Venus and Mars, occupy the outer edges of our solar system's habitable zone. However, with a surface temperature of 480 °C, Venus has traditionally been regarded as entirely hostile to all forms of life. In contrast, with an average surface temperature of −60 °C, Mars is too cold to support life on its surface. However, other possibilities remain. Photographs from the orbital missions of NASA and the European Space Agency (ESA) have shown that surface water on Mars was abundant in the past, leaving traces of gullies, creeks, streams, large rivers, lakes, and bedded sedimentary sequences. Even if they do not currently harbor life, they offer us more geochemical environments that could have once supported prebiotic chemistry in the past.

Mercury is the smallest planet in our solar system. It is too close to the Sun to harbor liquid water and life. Among the eight major planets, Earth is the only one with a single moon. Although Mercury and Venus have no moons, two dark, lumpy moons orbit Mars, namely, Phobos and Deimos, with the latter being the smallest moon in the solar system, only 11 km in diameter. Most likely, Mars captured them from the nearby asteroidal belt. However, Jupiter has 79 moons, Saturn has 82, Uranus has 27, and Neptune has 14. Even Pluto, which is now considered a dwarf planet, has five moons. So far, 219 moons of the planets are known in our solar system. Some of the frozen moons of Jupiter and Saturn have the potential to harbor microbial life.

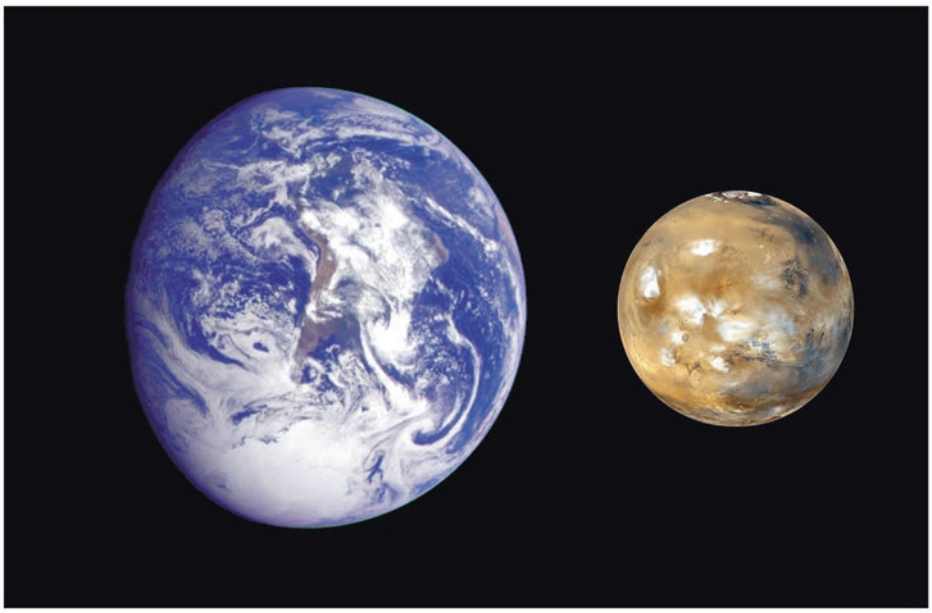
19.3.1 Venus

Venus is the second planet from the Sun and is the brightest object in the sky after the Sun and Moon. Often called Earth's twin, Venus has approximately the same size and density as Earth. Its surface shows evidence of extensive volcanism and lacks any evidence of plate tectonics. The planet has few impact craters. However, at around 480 °C, Venus has the hottest surface temperature in the solar system. The oven-like temperatures on Venus result from the greenhouse effect of crushing air pressure at its surface—more than 90 times that of Earth—as well as the planet's extremely dry, dense, and insulating atmosphere—composed of 96.5% carbon dioxide (CO_2), 3.5% nitrogen, and traces of other gases such as sulfur dioxide and sulfuric acid. At some point, a runaway greenhouse effect expelled whatever surface water it may have had to space. Venus at present is a vision of hell. Nevertheless, it is commonly believed that this planet once possessed a mild atmosphere and held sufficient water for Earth-style life to have flourished.

Moreover, despite its current toxic and hostile environment, some scientists have suggested that its upper atmosphere might still be hospitable and perhaps inhabited by microbes. Recently, the detection of the chemical phosphine (PH_3)—a biosignature signal in the cloud decks of Venus—has rekindled the possibility of some airborne microbes in its atmosphere [2]. Here on Earth, phosphine is produced by anaerobic microbial organisms, in some deep-sea worms, in the feces of badgers and penguins, and in our own intestines. The presence of phosphine in the oxidizing atmosphere of Venus is anomalous, especially because the gas has not been previously detected on any planet other than Earth. Some say that the only explanation for the chemical's source is something now alive. However, other scientists are more cautious about the implication of phosphine as an indicator of life. Recently reanalyzing their data, the original proponents have now been downgrading their claim. Phosphine levels are at least seven times lower than first claimed. This potential sign of life on Venus is fading fast.

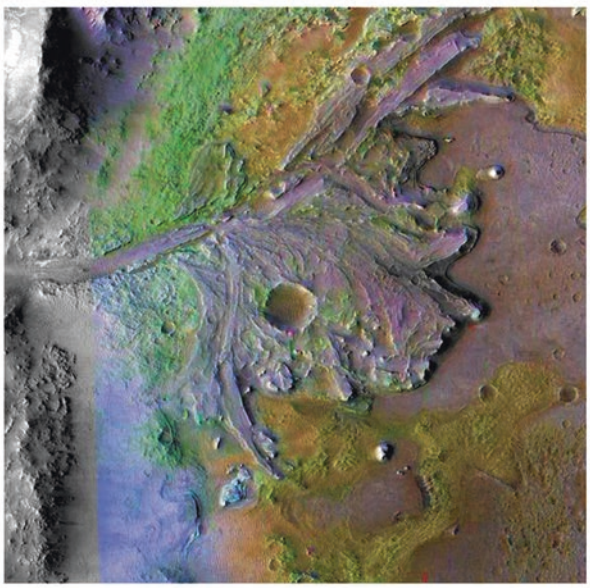
19.3.2 Mars

Like Earth, our sister planet Mars is a rocky planet, and like Earth's seafloor, much of the Martian surface is made of basalt. Mars is a cold, inhospitable desert today. Its average temperature ranges from −10 to −76 °C, with an average surface temperature of −65 °C and fluctuating temperatures from as high



A

B



C

Fig. 19.1 (**a**) Size differences between Mars (right) compared with Earth. Mars' diameter is approximately half that of Earth's, and its mass is about one-tenth that of Earth's. (**b**) The Gale crater (~154 km diameter) on Mars, a complex crater with a central peak, ~3.5–3.8 Ga old (Noachian), where the Curiosity Rover is currently exploring for life. (**c**) At the center of the photograph, Jezero crater (~50 km diameter) on Mars, ~3.5 Ga old (Hesperian), where Perseverance rover began to collect samples of rock and regolith for detection of microbial life. (Image credit: NASA)

as 25 °C to as low as −123 °C. The current Martian atmosphere is much different from that of Earth's, containing 95.3% CO_2, 2.7% N_2, and 0.1% O_2, as compared with Earth's atmospheric composition of 78.1% N_2, 21% O_2, and 0.04% CO_2. Its radius is about half that of Earth's, with a mass of about one-tenth that of Earth's. Consequently, Mars' gravity is less than half that of Earth's (Fig. 19.1a). Today, Mars is a bone-dry frozen desert, like the Dry Valleys in Antarctica. Part of its thin atmosphere freezes every winter to cover the poles with a white cap, but it provides little protection from heat loss

to the Martian surface. An average temperature of about −60 °C is too cold to support life on its surface. It appears that Mars lost nearly all its original atmosphere for a billion years, transforming its climate from one that might have supported life into a desiccated and frozen environment.

Nevertheless, when imagining where extraterrestrial life could potentially dwell, few places inspire the imagination like Mars. A quick start to life on Earth may indicate that life could emerge on other worlds in the solar system. Mars may well have harbored life in the past. Mars has long looked the most promising due to its proximity and similarities to Earth, its relatively pristine conditions, and its location at the outer edge of the habitable zone in our solar system. Due to the essential role of liquid water in the biochemical reactions that sustain living organisms, planetary bodies with global oceans or impact crater lakes are prime targets in searching for life beyond Earth. In addition to water, life would seem to require energy as a source of the essential building blocks. Because it fulfills these conditions by once having running water and a more robust atmosphere, Mars was capable of hosting ecosystems—and it might still be an incubator for microbial life today.

Based on their intersection relationships and the numbers of superimposed impact craters, the Martian geological time scale has been divided into four eons: the pre-Noachian (4.5–4.1 Ga), the Noachian (4.1 to about 3.7 Ga), the Hesperian (3.7–3.0 Ga), and the Amazonian (3.7 Ga to present) [3]. At the end of the Noachian, most geological activities such as rates of impact, valley formation, weathering, and erosion all slowed, but volcanism continued at a relatively high average rate throughout the Hesperian (Table 19.1). Evidence that liquid water used to be present suggests that Mars was once more habitable than it is today. During the Noachian period, about four billion years ago, when life was emerging on Earth, Mars enjoyed a warmer climate, with abundant liquid water on its surface, river valley networks, large crater lakes, and oceans. Many large impact craters scar the Noachian-aged surfaces. Dried riverbeds, impact crater lake deposits, polar ice caps, and minerals that form in the presence of water suggest that Mars had an environment that microorganisms could inhabit. Moreover, its subsurface may remain amenable as a habitat.

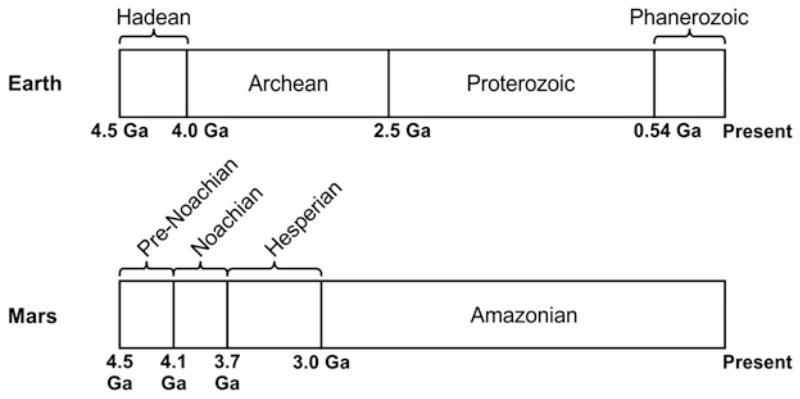
Searching for Life on Mars

Despite previous attempts to find it, we still do not know whether Mars has life or ever had life. The Viking landers of the 1970s conducted biological experiments to detect life on the Martian surface but failed to find any organic molecules. The scientific results did not demonstrate conclusive biosignatures at the two landing sites. To date, then, no proof has been found of past or present life on Mars. Nevertheless, Mars was not always a desolate wasteland. Cumulative evidence shows that Mars' surface environment had liquid water during the Noachian Eon (see Table 19.1) and may have been habitable for microorganisms. Dry riverbeds and minerals that form only with liquid water indicate that in deep time Mars had a thick atmosphere that retained enough heat for liquid water—a necessary ingredient for life—to flow on the surface. Like young Earth, it had vast oceans with clouds floating through the sky. Over the eons, the water was lost into space or locked underground. However, early conditions on a wetter planet could have been suitable for life to emerge. Mars remained warm and moist for millions or even billions of years, so microbial life might have had enough time to arise and proliferate. When conditions on the surface of Mars turned frozen and dry, life may then have become extinct there while leaving fossils behind. It is even possible, judging from some microbes on Earth that thrive miles underground, that extremophilic forms of early Martian life could have survived on Mars below its surface.

Three missions set out in the summer of 2020 on a journey to Mars, carrying large assortments of instruments to explore the red planet. The Hope orbiter, launched to Mars by the United Arab Emirates, was followed by the Chinese Tianwen-1, a combination of an orbiter, a lander, and a rover. NASA's latest Mars mission includes Perseverance, a 2200-pound rover, and Ingenuity, an experimental Mars helicopter. The Ingenuity helicopter was the first to attempt powered flight on another planet. The Perseverance rover's design is based on Curiosity, a successful NASA mission that arrived on Mars in 2012 to gather critical data. The car-sized Perseverance rover landed on Mars on 18 February 2021 on the floor of the Jezero crater and snapped a beautiful high-definition (HD) panorama of the landing site on Mars. It has been busy collecting samples from the Jezero crater and bringing them back to Earth to be studied for signs that life may have developed in this long-lived, ~50-km diameter impact crater lake. A river delta, point bars, and clay minerals have been detected around the crater. The Jezero crater is likely to have been habitable in the distant past (Fig. 19.1c). The crater formed around 3.5 million years ago within the large Isidis impact basin when microbial life was present on Earth.

If we are fortunate enough to find even one sample containing a microbe, a living organism, or a fossil on Mars, then we will have identified the wonderful circumstance of two sister planets both supporting life in the same early epoch. In that case, did life arise independently on these two sister planets, only to be wiped out on Mars when the climate irrevocably altered? Or might there be subsurface refugees on Mars, i.e., some still lingering life-forms? It is reasonable to believe that life also appeared on Mars under similar conditions in hydrothermal crater lakes, with similar cosmic ingredients, and at about the same time. Perhaps life started simultaneously on Earth and Mars around four billion years ago, when Mars had a much thicker atmosphere, warmer

Table 19.1 Earth and Mars compared, showing the major factors that affect their current habitability

	Earth	Mars
1. Average distance to Sun	1 AU*	1.5 AU*
2. Average surface temperature (°C)	15	-56
3. Greenhouse effect (°C)	33	7
4. Atmospheric pressure (bar)	1	0.006
5. Liquid water	Abundant	Frozen
6. Main gases in the atmosphere	78% N_2, 21% O_2	95.3% CO_2, 2.7% N_2
7. Diameter (equatorial)	12,756 km	6794 km
8. Mass	5.974×10^{24} kg	6.418×10^{23} kg
9. Escape speed	11.2 km/s	5 km/s
10. Geological time scale (Ga)		

*Astronomical Unit (AU) = Earth-Sun distance, about 150 million km

climate, and surface water. The evidence that Mars was once a wet planet in the Noachian is incontrovertible due to the sedimentary structures left by lakes, flowing rivers, and large oceans. Perhaps, as the widespread cratered surface and volcanic rocks testify, early Mars was warmed in the past by volcanic emissions and asteroidal impacts. Unlike Earth, however, Mars did not develop plate tectonics, so its surface has faithfully retained its early history. It provides us with a rare glimpse of how early Earth might have looked.

If Mars ever harbored life during the Noachian Eon, when liquid water flowed freely on its surface, then some microbes stowed away in the nooks and crannies of the rocks could have been ejected from Mars by impacts and, like the ALH84001 Martian asteroid found in Antarctica, traveled to Earth for free. We know that some microbes are incredibly hardy and may be able to survive high doses of solar radiation for long periods during an interplanetary journey after being blasted off their home by an asteroidal impact. Some believe that Earth life originated on Mars and was brought to this planet in this way, aboard a meteorite. Orbital dynamics shows that it is much easier for rocks to travel from Mars to Earth than the other way around. If Mars was ever a living planet, it died early and lost its atmosphere. However, to date, no indigenous Martian organisms have been discovered to support this exciting possibility of interplanetary travel of early life.

We do not know for sure how Mars lost its atmosphere and oceans. Mars was certainly habitable in the Noachian when both Mars and Earth were covered with two protective shields—a relatively thick atmosphere and a strong magnetic field. Earth has retained both, but Mars lost them billions of years ago and so became more vulnerable to cosmic assaults. Mars' loss of its magnetic field strongly affected surface environments through increased radiation, significantly degrading surface habitability. Because of its small size, lower mass, and lesser overall gravity, Mars' internal heat was also lost quickly, ceasing widespread volcanism. Without volcanism, Mars could not recycle atmospheric carbon dioxide. Moreover, part of its early atmosphere was lost during the early bombardment period. Because of the planet's low escape velocity, the Martian atmosphere was particularly prone to impact erosion. With its weak atmospheric pressure, Mars could not keep its water in liquid form.

Nevertheless, if microbial life did exist in the Noachian sediments, it may have been extremophiles that could withstand solar and cosmic radiation. Is it possible that biosynthesis began separately but simultaneously on both nearby planets? If so, what genetic code did it use? Was Martian life genetically coded with nucleic acids? Did it use proteins for enzymatic catalysis? Did it have left-handed amino acids and right-handed sugars? Irrespective of the answers to these questions, should we discover life on Mars, we will have the first opportunity to compare the biology of two different planets. Or was life transported from one world to another by asteroidal impacts, circumventing the need for independent origins? The optimistic conviction that we may have all these answers within a few decades lends an enormous impetus to the search for signs of life, past or present, on the red planet.

Evidence obtained by NASA's Curiosity rover from Aeolis Palus in the Gale crater on Mars strongly suggests an ancient hydrothermal freshwater lake, which could have been an environment hospitable to microbial life [4]. The Gale crater was created by an asteroidal impact in the Martian Noachian–Hesperian transition, around 3.8–3.6 Ga, roughly contemporaneous with Eoarchean Earth (Fig. 19.1b). Therefore, Curiosity's exploration of the Gale crater for signs of life is highly exciting. Curiosity detected mudstones deposited from water containing the CHONPS elements. It has also found some structures, akin to stromatolites, which could have been microbially generated. Additionally, drill samples from the Gale crater data provide new evidence for organic molecules—theophanic, aromatic, and aliphatic compounds—in the 3.5-billion-year-old sedimentary rocks. These organic molecules are important because they tell us about the chemical pathways of their formation and preservation. Curiosity data indicate that several billion years ago, a hydrothermal lake inside the Gale crater might have held all the ingredients necessary for life, including chemical building blocks, energy sources, and liquid water. These recent discoveries of possible biotic activity have been tantalizing.

Curiosity has shown that the Gale crater was habitable around 3.5 billion years ago during Hesperian [5], with conditions comparable to those on early Earth, where microbial life evolved (Fig. 19.1b). The hope of discovering contemporary life on Mars increased in December 2014 when the Curiosity rover recorded intriguing bursts of methane. Specifically, Curiosity found a 10-fold spike in atmospheric methane and detected other organic molecules in a powdered rock sample collected with the robotic laboratory's drill [4]. Around 90% of the methane on Earth comes from the metabolism of organisms, so it seems plausible that life on Mars is also emitting this gas. Because methane is an unstable gas, its enduring presence indicates an active source on the planet to maintain such levels in the atmosphere. Moreover, background levels of methane in Mars' atmosphere show substantial seasonal variations. The presence of methane may indicate the existence of methanogens, such as archaea, which produce methane as a metabolic byproduct under anoxic conditions. Localized methane sources released from surface or subsurface reservoirs may indicate hyperthermophilic activity [4]. The detection of organic molecules and methane on Mars has far-ranging implications regarding potential past life on Mars. However, no firm conclusions can be drawn until such Martian rocks can be brought back to Earth for laboratory confirmation. Therefore, the Perseverance mission of sampling the rocks of the Jezero crater and bringing them back to Earth is highly exciting. Perseverance is playing an extremely critical role in our understanding of our place in the universe.

One major outstanding question crucial for biosynthesis concerns the longevity of liquid water on Mars. During its exploration of the Gale crater, the Curiosity rover also discovered the sediments of an ancient lakebed, suggesting that water was present for at least a few million years. In the upper part of the ancient lakebed sequence, there is a high concentration of sulfate salts from the late Hesperian Eon starting around 3.7 Ga, a time when Mars had active volcanoes and plenty of water in its shallow seas and lakes. The high concentration of salts in the upper sequence points to a period of rapid evaporation at the end of the Hesperian (~3.4 Ga) when Mars began the transition to the cold desert we know today. The saline impact crater lakes on Mars endured a period of atmospheric loss and may have been some of the last surface waters on the red planet. Porous rocks on the Martian surface are typical of salty lake deposits, such as magnesium sulfates, bromides, and chlorides. Some scientists speculate that iron oxide balls strewn on the Martian surface resemble those formed by underground bacteria found on Utah's sandstone hills. Perhaps these iron balls were also formed by ancient Martian microbes.

The surface of Mars is now dry and cold, like Antarctica. Could Mars host life now? A recent discovery of three buried lakes of salty water under Mars' icy surface has sustained hope for the survival of subsurface life. If there is any life on Mars now, it needs liquid water. Most of its liquid water is locked away in subterranean reservoirs. Perhaps, like frozen lakes in Antarctica, life may be hidden away in those underground lakes. Furthermore, because a diversity of extremophiles is known to survive in the most inhospitable habitats on Earth, it is likely that Martian extremophiles could still be hanging on.

Martian Meteorites

More than 270 asteroids collected on Earth, which were ejected from other planets by impact events, have been traced back to Mars. Martian meteorites formed millions of years ago as asteroids crashed onto the Martian surface and knocked Martian rocks off the planet, and some of these

eventually fell onto Earth. The knowledge that these strange space rocks came from Mars was obtained when data on Martian rocks and atmosphere coming from NASA's orbiting spacecraft, surface landers, and rovers were analyzed in the 1980s. These meteorites have elemental and isotopic compositions that match rocks and atmospheric gases on Mars. Another fingerprint of Martian meteorites is the unique triple oxygen isotope (^{16}O, ^{17}O, ^{18}O) composition. These rocks could only have come from Mars. It turns out that about 50 kg of Martian meteorites fall to Earth each year, mainly in the oceans [6].

One evening, while I was doing fieldwork in 1984–1985 on the Allan Hills of Victoria Land in Antarctica, a few NASA scientists in a snowmobile visited our camp and showed us some meteorites they had collected that day from the ice field. These meteorites looked like black, burnt rocks. Little did I know at that time that one of the meteorites collected from the Allan Hills that year would draw worldwide attention. This was Allan Hills 84001 (ALH84001), originally formed 4.5 Ga on Mars, then blasted from its surface by the impact of another meteor about 17 Ma, landing on Antarctica about 13,000 years ago. In 1996, a team led by David McKay of NASA's Johnson Space Center in Houston made headlines around the world by announcing the discovery of microbial fossils recovered from ALH84001. President Bill Clinton went on TV to mark the event. According to the interpretation of McKay's team, the structures found on ALH84001 contain elliptical rope-like 'bacterial fossils' [7]. These are 20–100 nm in diameter, similar in size to theoretical nanobacteria but smaller than any known cellular life. Some of the tiny grains of magnetite found along with the putative fossils also appeared to be biogenic.

These claims of Martian life were subsequently challenged, and the putative microfossils were ruled out as life because of their extremely small size. However, this could still be evidence of protocells—a precursor to life. At the least, Martian meteorites, such as ALH84001 that fell in Antarctica, Nakhla in Egypt, and Shergotty in India, contain features suggestive of ancient biogenic activity on Mars [8]. The morphological similarities between terrestrial microfossils, biofilms, and the three Martian meteorites' features are intriguing but so far inconclusive. Still, if rocks from Mars can land on Earth, perhaps enclosed microorganisms can too. Despite it all, enthusiasm for finding life on Mars remains unabated.

Colonization of Mars

Mars has captivated humans since we first set eyes on this red planet in the night sky. Mars has always been an enchanting focus of scientific study and possible human colonization. The surface conditions and water presence in the past make it arguably the most hospitable of the other planets in the solar system. It is certainly the most explored planet in the solar system other than Earth. Despite its small size, the planet's land area is roughly equivalent to the surface of Earth's continent. Once every 26 months, Earth and Mars are aligned in way that minimizes travel time and expenses, enabling spacecraft to make the interplanetary journey in roughly 6 months. With improved space technology, travel time would be reduced considerably. SpaceX CEO Elon Musk has proclaimed that humanity must become a 'multiplanetary species' if we are to survive the climatic calamities and environmental disasters caused by us. He envisions a million people living on Mars before the end of this century, our second home in the solar system. Some scientists believe that human colonization of the harsh and exotic atmosphere on Mars might accelerate our species' evolution.

As epitomized by Kim Stanley Robinson's Mars trilogy, a settlement on Mars is one of science fiction's prominent themes. Beyond sample-return and human-crewed missions, we will establish a settlement there as technology improves. However, Mars presents a brutal environment for human habitation. With surface temperatures becoming low as −153 °C at the poles and an atmosphere too thin to retain solar heat, this bitterly cold and desiccated world entirely lacks oxygen. A Mars mission would be a phenomenal undertaking by any standards, with monumental risks and challenges. Nevertheless, for NASA, it is their next destination. In 2015, President Obama wrote an 'op-ed' at CNN.com describing the ambitious goal of humans to reach Mars in the next two to three decades:

> We will have set a clear goal vital to the next chapter of America's story in space: sending humans to Mars by the 2030s and returning them safely to Earth, with the ultimate ambition to one day remains there for an extended time.

NASA has now laid out detailed plans to get there by 2030, proposing the Orion Multi-Purpose Crew Vehicle (MPCV) and Space Launch System (SLS). NASA's mission is not the only proposal to send humans to Mars. There are also ambitious private projects, including 'SpaceX' by Elon Musk and Jeff Bezos's New Glenn and New Armstrong reusable launch vehicles, which may expedite, if not lead, the crewed missions. Mars exploration may follow the model of the investigation of Antarctica, establishing a permanent research base in the initial stage, with the potential for later expansion. Natural caves and lava tubes may provide shelter from radiation and micrometeorites on the Martian surface. Geothermal energy is also believed to be present in the equatorial region. Establishing a permanent settlement would be extremely complex but less complicated, requiring less infrastructure to be shipped to Mars than a return mission. For instance, photosynthetic bacteria and plants could convert CO_2 in the atmosphere to breathable oxygen and edible carbohydrates. Extensive human settlement on Mars will eventually depend on the development of geoengineering technologies.

19.3.3 The Icy Moons of Jupiter

Beyond Mars and the asteroidal belt lies the giant Jupiter, the largest planet in the solar system. Jupiter has 79 confirmed moons, with the 4 largest ones being the most well-known. These four moons are likely to have been first observed by Galileo Galilei in 1610; thus, they are often called the 'Galilean moons.' Spheroidal and among the largest moons in the solar system, Io, Europa, Ganymede, and Callisto would be considered dwarf planets (like Ceres) if they orbited the Sun. The largest, Ganymede, is bigger than Mercury! Jupiter's icy moons—Europa, Ganymede, and Callisto—may possess energy sources, biogenic molecules, and vast liquid oceans beneath their icy crusts, evoking the possibility of them being abodes for past or present life. All these icy moons show evidence of impact cratering. NASA has proposed an ambitious mission to send a spacecraft to study these three moons—the Jupiter Icy Moons Orbiter (JIMO).

Europa

Europa is estimated to be about 4.5 billion years old, about the same age as Jupiter. It is approximately 3100 km in diameter, smaller than Earth's moon but larger than Pluto. Its average distance from the Sun is about 780 million km. Europa's icy surface temperature never rises above −160 °C. Its icy crust is highly reflective, giving Europa its bright surface. Like our Moon, Europa is tidally locked such that the same side faces Jupiter (Fig. 19.2a). Data from the Galileo spacecraft suggest that Europa is composed of silicate rock, with an iron core and a rocky mantel much like Earth. Unlike Earth, however, Europa's rocky interior is surrounded by a layer of water and/or ice (~170 km), covered by an ice shell perhaps 15–25-km thick, floating on an ocean 60–150-km deep.

However, the salty ocean of liquid water that lies beneath Europa's cratered ice crust may provide a habitat capable of sustaining life. Like Earth, Europa presumably received organic compounds from meteorite impacts. As we know, the possibility of life depends on the availability of liquid water, cosmic building blocks, and energy sources. The impact cratering suggests that Europa had a good supply of cosmic biomolecules. There has been much speculation that Europa's interior ocean is heated by hydrothermal vents on the deep ocean floor. Such ocean vents could produce sufficient chemical energy to sustain life. The case for life on Europa was significantly enhanced in 1995 when the operators of the Hubble Space Telescope determined that there may be enough oxygen in that ocean to support life. Unlike the oxygen in Earth's atmosphere, however, that on Europa is not of biological origin. The oxygen formed on Europa may have arisen when energetically charged particles from the Sun were trapped by Jupiter's powerful magnetic field and crashed onto the moon's icy surface with sufficient energy to split its water into oxygen and hydrogen. A key question is whether enough oxygen reaches the ocean to support the oxygen-based metabolic processes of aerobic bacteria.

Ganymede

Another Galilean satellite of Jupiter, Ganymede, is the largest moon in the solar system. Its surface consists of bright and dark terrain. The bright terrain is dominated by water, and the dark landscape is silicate-rich rock. Ganymede is layered with a Fe or Fe–S core, a silicate mantle, and an outer ice crust. The icy surface is highly cratered, faulted, and grooved with a network of interconnected rectilinear ridges known as reticulate terrain.

Callisto

The second largest of the Galilean moons, Callisto, is almost as big as Mercury. It may also have a liquid ocean beneath its crust. Impact craters dominate Calisto's surface, which is otherwise relatively smooth and covered with dark material that seems to fill in topographic lows and small craters. The darker areas likely contain silicates, possibly organics; ice dominates the brighter regions.

19.3.4 Saturn's Moons

Saturn is well known as a gas giant with a powerful ring system. It also has 82 moons. Saturn's moons range from larger than the planet Mercury—the giant moon Titan—to as small as frozen Mimas, about 396 km in diameter. NASA's Cassini spacecraft discovered many new moons of Saturn and provided critical information about them. While the larger moons are spherical, others are shaped like a sweet potato (Prometheus), a regular potato (Pandora), a meatball (Janus), and even a sponge (Hyperion). Among these moons, Enceladus and Titan are highly promising for harboring life.

Enceladus

Enceladus is the sixth-largest moon of Saturn, about 500 km in diameter. With geothermal activity, water vapor, and possibly under-ice oceans heated by tidal effects, this frozen moon might be hospitable to alien life. NASA's Cassini spacecraft has provided intriguing images of icy jets spewing from a suspected underground liquid ocean. Enceladus has a rocky core, an icy shell, and a salty underground sea. With a pH of 11–12, Enceladus's subsurface ocean is quite basic, like soda lakes on Earth, such as the Lonar Lake in India, which hosts a variety of life. It may also have organic compounds and a source of energy derived from the friction created by its orbit around Saturn.

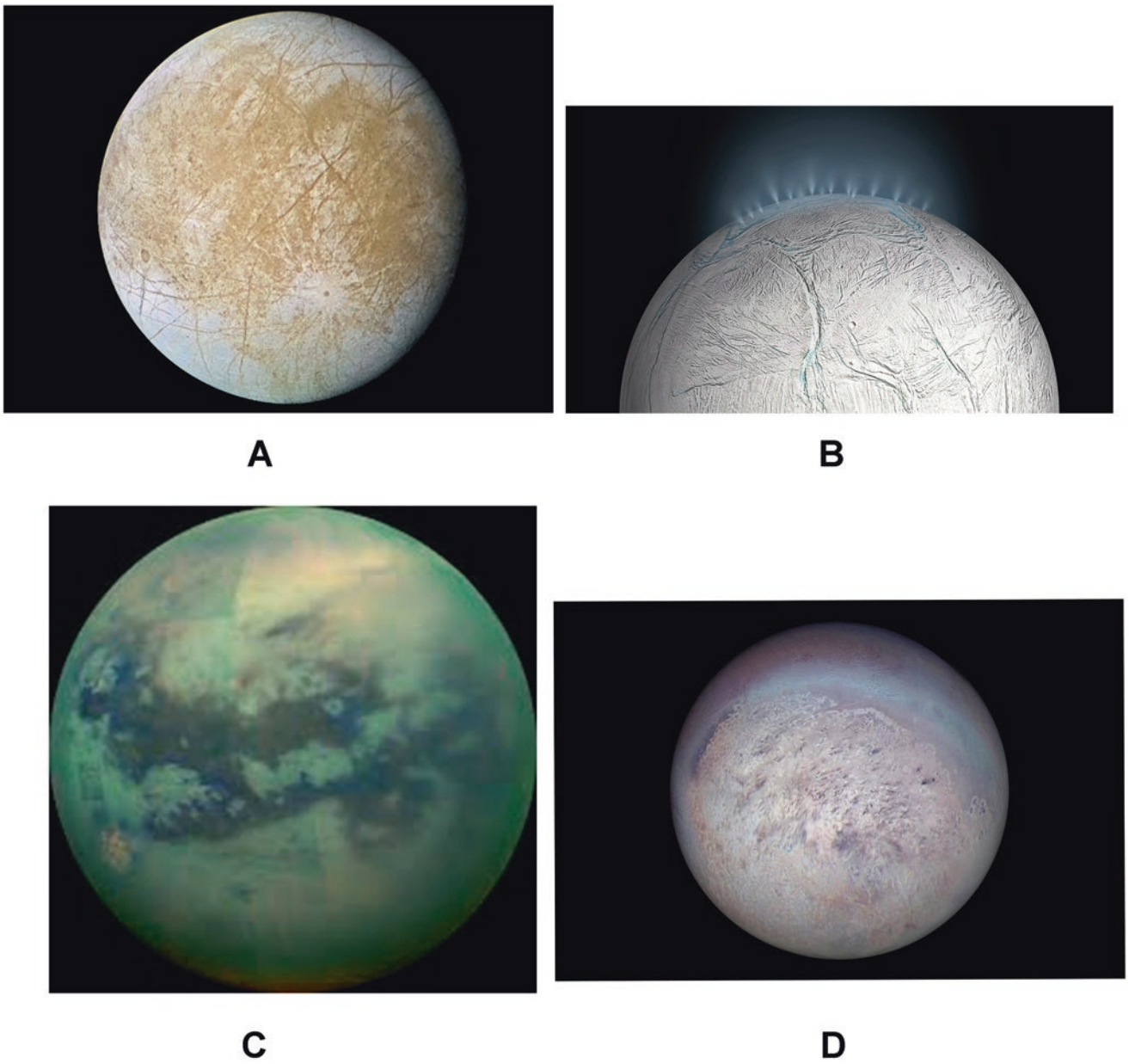


Fig. 19.2 Icy moons in the solar system where life might exist. (**a**) Europa, Jupiter's moon. (**b**) Enceladus, Saturn's moon, showing multiple geysers. (**c**) Titan, another one of Saturn's moons. (**d**) Triton, Neptune's moon. (Image credit, NASA)

Its more than 90 geysers spew plumes of salty water vapor, icy particles, organic compounds, and hydrogen molecules from its underground ocean (Fig. 19.2b). The most plausible source of this hydrogen is ongoing hydrothermal reactions of hot water flowing through the cracks of the seafloor's rocky core, a process called 'serpentinization.' It has been postulated that silicate controls the pH, salinity, and abundance of silica in the Enceladus ocean. The geysers contain surprising ratios of molecular hydrogen, carbon dioxide, and methane, all extremely far from equilibrium. The presence of hydrogen may indicate unique underwater hydrothermal vents, a possible habitat for hyperthermophiles [9]. Any microbes that exist there may have access to two different sources of metabolism-supporting energy—molecular hydrogen and the heat provided by the hydrothermal vents. Its global ocean, unique chemistry, and internal heat make Enceladus a promising lead in searching for life in the solar system.

Titan

Titan is the largest of Saturn's moons and probably the most enigmatic moon in the solar system. It is as large as the planet Mercury but far colder than Earth. Its surface lacks stable liquid water but contains liquid methane and ethane.

Exerting 50% more surface pressure than Earth's, its thick atmosphere is chemically active and rich in methane and nitrogen, possibly produced by oceanic microbes (Fig. 19.2c). Recently, organic molecules have been detected in its atmosphere, including benzene (C_6H_6) and cyclopropenylidene (C_3H_2). Models of Titan's thermal history indicate the presence of an ocean as much as 200 km deep, consisting of an ammonia–water solution, beneath a crust of water ice. This could have been a suitable habitat early in the moon's history for the development of life forms such as methanogens. Findings from NASA's Cassini spacecraft show that hydrogen disappears near Titan's surface, thus suggesting that Titan might have harbored methane-based life. The most intriguing aspect of Titan is the presence of abundant methane itself. Atmospheric methane does not usually persist when hit by sunlight; it breaks down into hydrogen atoms and lighter molecules, which readily combine with others to form more extended and more complex organic compounds, such as acetylene and ethane. Microorganisms catalyzing methane from acetylene and ethane could represent a significant source of methane in Titan's atmosphere. Microbial life on Titan could breathe hydrogen and consume acetylene, producing methane as a byproduct. However, many consider life on Titan life unlikely because of the absence of surface liquid water. Whether life exists in subsurface oceans on Titan, Enceladus, and Europa is a question for the next generation of explorers to answer.

19.3.5 Neptune's Moons

Neptune has 14 known moons and 6 known narrow rings. Since the planet's name is derived from the Roman god of the sea, Neptune's moons are named after various lesser sea gods and nymphs of Greek mythology.

Triton

Neptune's largest moon, Triton, is unusual as the only body in the solar system that orbits in the opposite direction than its planet's rotation—what is termed a 'retrograde orbit.' Triton's retrograde orbit suggests that it is a captured object originally formed, like Pluto, as an independent icy planetesimal from the outer solar system's Kuiper Belt, a disk-like region occurring within the solar system plane beyond the orbit of Neptune, at a distance of 30.1 astronomical units (AU) (1 AU being the mean distance from Earth to the Sun), consisting of icy bodies left over from solar system formation. It is believed to be populated by hundreds of thousands of rocky, icy bodies, each larger than 100 km across, along with a trillion or more comets. Triton and Pluto are both Kuiper Belt objects (KBOs). With a diameter of 2700 km, Triton is like Pluto in size, density, and surface composition (Fig. 19.2d).

Voyager 2 revealed fascinating details about Triton: its surface temperature is 38 K (−235 °C), and it has active geysers. Triton is layered and consists of a crust of frozen nitrogen overlying an icy mantle with a core of rock and metal. Its atmosphere is composed of nitrogen, methane, and carbon monoxide. Ice volcanoes spout what is probably a mixture of liquid nitrogen, methane, and dust, which instantly freezes and then snows back down to the surface. Icy vents, fissures, and lavas indicate that Triton is geologically active, whereas crater counts on the frozen crust suggest that resurfacing took place as recently as 10 Ma.

Early Triton may have had an ocean and been more habitable before its capture. After its capture, tidal heating from Neptune might have melted a vast subsurface ocean and caused the crust to split. The discovery of geysers spouting water from Enceladus and Europa is prompting another look at Triton's plumes. Like Enceladus and Europa, Triton's water plumes may come from its interior ocean. Water in Triton's ocean could be much colder than the usual freezing point, but ammonia could preserve it in the liquid state. If Triton has kept some internal heat to support a liquid ocean underneath its icy shell, then this could be a potentially habitable site in the solar system. If the future Trident mission of NASA confirms that an ocean exists on Triton, then an even broader expanse of the solar system may be capable of sustaining life.

19.3.6 Pluto

Once considered the ninth and most distant planet from the Sun, Pluto was stripped of its status as a planet in 2006 by the International Astronomical Union and reclassified as a dwarf planet. It orbits within the Kuiper Belt. It is now the largest known dwarf planet in the solar system. It is also one of the largest known KBOs (Kuiper Belt objects) at about 2370 km in diameter, less than one-fifth of Earth's diameter, and only about two-thirds as wide as Earth's Moon. This fascinating world has blue skies, spinning moons, mountains, valleys, plains, and a heart-shaped glacier bigger than Texas, and it snows—but the snow is red. Pluto is probably composed of a mixture of 70% rock and 30% water ice. Its surface mainly consists of frozen nitrogen, methane, and carbon monoxide ices, with a thin nitrogen atmosphere of 8–15 μbar surface pressure. The dwarf planet also has polar caps and regions of methane and nitrogen. It probably has a mantle of ice overlying a rocky core.

Pluto has five known moons, Hydra, Nix, Charon, Kerberos, and Styx, of which Charon is the largest, about half the size of Pluto itself. Like regions of Pluto, much of Charon's surface is free of craters—suggesting that the surface is quite young and geologically active. In 2006, NASA launched its New Horizons spacecraft on a 6-month reconnaissance flyby study of Pluto that revealed a variety of

surface features, including mountains such as Tenzing Montes that reach as high as 3500 m, comparable to the Rocky Mountains. Perhaps these mountains are formed on a bedrock of water ice over a subsurface ocean. NASA's Hubble telescope has revealed evidence that Pluto's crust contains complex organic molecules. On average, Pluto's temperature is −232 °C, making it too cold to sustain life. However, Pluto's interior may harbor a warmer ocean deep inside.

19.3.7 Exploration of Asteroids

Asteroids—small rocky bodies leftover from the accretion process in the disk around the very early Sun that did not become parts of planets and moons—are as ancient as the solar system, some 4.5 billion years old. Millions of them populate the asteroidal belt between Mars and Jupiter. As discussed earlier, while primarily considered a threat from above, asteroids are also life-givers. The water they bring is a crucial ingredient for life as we know it. Carbonaceous chondrites—rich in the building blocks of life and water molecules—might have delivered the earliest building blocks of life in the solar system. Both their age and makeup have made asteroids increasingly attractive to astrobiologists, spurring a recent suite of missions to asteroids by NASA, the Japanese Aerospace Agency, the European Agency, the Russian Space Agency, and the China National Space Administration.

Ceres

At 2.8 AU from the Sun and some 950 km in diameter, the dwarf planet Ceres is the largest planetesimal in the main asteroidal belt. Termed an 'asteroid' for many years but much bigger than its rocky neighbors, in 2006, Ceres was reclassified as a dwarf planet. Unlike other asteroids, Ceres is spherical like other terrestrial planets (Mercury, Venus, Earths, and Mars) while it is also less dense than those planets. Its surface temperatures can rise as high as −38 °C. It has a layered interior, a solid core, a mantle made of water ice, and a surface that contains clay and ammonia-bearing hydrated minerals, magnesium sulfate, water ice, carbonates, salt, and organic materials. NASA's Dawn mission made the first unambiguous detection of organic molecules on an orbiting asteroid [10]. Analyzing the light reflected off Ceres from the orbit, the Dawn space probe was able to determine the types of molecular matter on its surface. Near the large (~50 km) Ernutet crater in Ceres' northern hemisphere, the space probe found the distinctive molecular signature of a class of organic molecules with methyl and methylene groups. These are chemical chunks of carbon and hydrogen atoms grouped in aliphatic chains, such as CH_3 and CH_2. Ceres may also retain internal heat from its formation period along with a subsurface ocean where microbial life could have developed. This dwarf planet joins Mars, Europa, and Enceladus on the list of locations in the solar system that may contain life.

Bennu

The Stardust mission of NASA has discovered the building blocks of life in comets and interstellar dust, supporting the idea that those first life-building molecules are prevalent in the cosmos. Such wide distribution of cosmic ingredients for life strengthens the argument that life in the universe may be more common than rare. If meteorite bombardments seeded young Earth with raw materials, let us start searching for life's essential ingredients at the source—with near-Earth objects (NEOs) that orbit about 300,000 km above Earth. Numbering more than 14,000, NEOs are the asteroids closest to Earth. Bennu, a spheroidal carbonaceous asteroid about 500 m in diameter, is under extensive observation by the defunct Arecibo Observatory as a potentially hazardous object. It orbits the Sun every 436 days, coming extremely close to Earth, within 0.002 AU, every 6 years. Bennu is a rare B-type asteroid (primitive and carbon-rich), expected to have organic compounds and water-bearing minerals such as clays. Bennu's size, composition, proximity, and threat to Earth make it one of the most fascinating and accessible NEOs.

A NASA spacecraft of the OSIRIS-REx mission successfully touched down on Bennu in October 2020 and will return samples to Earth for analysis in 2023. Studying samples from an asteroid like Bennu should provide astrobiological insights into the early solar system, our cosmic connections, and even the origin of life. The knowledge of organic compounds in the solar system is the primary reason for investigating these asteroids. Did they deliver organic material and water to the surface of our planet four billion years ago that primed the emergence of life? The Murchison meteorite has already provided a wealth of such information. It is likely that Bennu, too, carries organic material from the young solar system and will provide pristine evidence of a wide range of organic compounds essential to life.

19.3.8 The Habitability of Icy Ocean Worlds

The icy moons of Jupiter, Saturn, Uranus, and Neptune are mainly built from frozen water, which becomes hard rock at frigid temperatures far from the Sun. However, even a tiny amount of internal heat can turn that ice into liquid water. Exploration of icy ocean worlds in the solar system offers the chance to anticipate what features of habitability to look for on watery exoplanets. If a watery world is geologically active enough, and if meteorites deliver organic compounds, then the distance from their suns may be less of an issue than it would be otherwise. The cratered surface of these

frozen moons suggests that meteorite impacts delivering critical ingredients of life are common. Life may have started around hydrothermal vents here on Earth, where hot fluids rise from the magma, creating complex reactions with water and cosmic ingredients. It may be that hydrothermal vents are also actively spewing nutrients and energy into the dark oceanic depths of Enceladus. Similarly, the base of Europa's ocean may feature hydrothermal activity, making it a suitable place for life to develop and thrive. If comparable vents are present in other watery worlds along with life-building organic compounds, then life may have started there as well.

In summary, alien life may be lurking right in Earth's cosmic backyard. Many believe that the first step to finding alien life in the broader universe should be exploring the icy water moons of the solar system for past or present life. The two asteroids Ceres and Bennu, the frozen moons of Jupiter and Saturn, and Kuiper Belt objects such as Pluto all show that life might exist in unlikely places outside the habitability zone in our solar system. Mars, Europa, Enceladus, and Titan offer the best hopes for finding water or other solvents within which organic molecules might move toward biogenesis and thus discovering extraterrestrial life, either alive or in the fossil form. Although most astrobiologists limit their hopes of finding microbial life to just one or two of them, given that so many objects in this solar system potentially harbor microbial life, billions of such possibilities surely exist throughout the galaxy. Just finding microbial life on a planet or moon in our solar system would be the first sign that we are not alone in the universe. This would change our view of the cosmos forever.

19.4 Searching for Life in the Milky Way

Perhaps the most promising place to find life beyond Earth is not our solar system at all. Our Sun is but one among the 100 billion stars in the Milky Way, which is a spiral galaxy well over 100,000 light-years across. Our solar system is positioned on a spiral arm 30,000 light-years from the center. From Earth's outpost, we have begun to peer across the void. We can already dimly make out the light from planets orbiting distant stars. Many other planets in the Milky Way must occupy habitable zones. To support life, planets must orbit in the right place to have liquid water, not too close and not too far from their star. Similarly, life will not emerge or survive for long near the center of their galaxies, where the high density of stars means that several supernovae could be exploding at any given time. Planets located at the outer edges of the spiral arms, far from the galactic center, face fewer hazards and are therefore more hospitable. If we find lots of planets like ours, then we will have many reasons to believe that we are not alone.

19.4.1 Exoplanets

One of Earth's most basic life-supporting attributes is indeed its location and size, its seemingly ideal distance from the Sun, keeping water in its liquid state, and enough gravity to hold the atmosphere down. In any planetary system in the galaxy, there are regions at specific distances from the central star where a surface environment like Earth could occur—what we have called the habitable zone. Astronomers have now detected many of these exoplanets in our galaxy, located in habitable zones that may have suitable conditions for life, supporting the notion of alien life on other worlds. In 2016, astronomers announced an exoplanet orbiting in the habitable zone of Proxima Centauri—at just 4.2 light-years from our solar system, the star closest to the Sun. Known as Proxima b, the newly discovered planet is about 1.3 times more massive than Earth, suggesting a rocky rather than gaseous world.

The first exoplanet was discovered in 1995, but their number has exceeded in recent years. Hidden by the bright glare of the stars they orbit, exoplanets are extremely hard to see directly with telescopes. Exoplanets are detected using transit photometry and Doppler spectroscopy. In 2009, NASA launched a spacecraft called Kepler space telescope to look for exoplanets, especially Earth-sized worlds in a Sun-like star's habitable zone. By monitoring more than 150,000 stars using the transit method, Kepler has discovered 2500 confirmed exoplanets and estimated the possibility of perhaps 2245. However, the most powerful method to obtain information about a faraway planet is using spectrometry to study atmospheric composition for surface qualities and biosignatures, signs of life.

A recent statistical study has estimated that there are a trillion exoplanets in our galaxy alone. According to a recent tally by NASA, there are more than 5000 known exoplanets, but the majority are big worlds like Jupiter and Neptune. Of these, 1264 are the so-called 'ice giants,' 1043 are 'gas giants,' and 781 are 'super-Earths.' A super-Earth is a larger and more massive planet than Earth, although still made of rock—perhaps with continents and oceans—and an atmosphere. Only about 160 exoplanets are Earth-sized or smaller, and, of these, 50 or so may have the potential to support life as we know it. When the NASA space telescope called 'TRAPPIST' was trained on the ultracool dwarf star known as Trappist-1, which is located about 39 light-years from Earth, seven Earth-sized temperate rocky planets were discovered orbiting this star. This was the first time that so many worlds of this kind have been found circling the same star. Three of those planets are located within the habitable zone, increasing the probability that liquid water flows on their surfaces. Could any of these planets harbor life? We do not yet know. However, serious studies are in progress to determine whether their atmospheres could support life.

19.5 Search for Extraterrestrial Intelligence (SETI)

An alternative approach to finding life on exoplanets is to hypothesize that technologically advanced alien civilizations also exist in our galaxy. If far-off worlds harbor such intelligent life, we can try to make contact by searching for their communications, or more aptly, for 'technosignatures' [6]. This is the logic behind the search for extraterrestrial intelligence (SETI). The first serious attempt to contact alien civilizations began in 1960 with Project Ozma, when astronomer Frank Drake looked through a radio telescope at two Sun-like stars located 11 light-years away, hoping to pick up a signal of intelligent life. The first popular and accurate book in this field was *Intelligent Life in the Universe* by two world-renowned astronomers, I.S. Shklovskii and Carl Sagan [11]. Their ideas sparked global interest in the search for extraterrestrial intelligence (SETI). Their book has inspired millions of people, especially science fiction enthusiasts. Such science fiction illustrate our creative yearnings to explore humanity's future within our cosmic envelope as well as all the hopes and fears that come with that heroic quest. Do we have cosmic siblings? Is there an interstellar family of intelligent beings who also look up to the stars with questions? Or are we alone, the universe's only child? As Carl Sagan reminded us: 'The universe is a pretty big place. If it's just us, seems like an awful waste of space.'

In 1961, Drake wrote an equation to quantify the likelihood of finding a technologically advanced alien civilization. The so-called Drake equation took into account factors such as the fraction of stars with planets around them and the fraction of these planets that would be hospitable to life to estimate the number of civilizations in the Milky Way that might seek to communicate beyond their own worlds. The equation is shown below [6]:

$$N = R^* \times f_p \times n_e \times f_l \times f_i \times f_c \times L$$

Its seven variables are defined as follows:

N = the number of civilizations in the Milky Way with which communications might be possible.
R^* = the rate of formation of stars suitable for the development of intelligent life.
f_p = the fraction of those stars with planetary systems.
n_e = the number of planets, per solar system, with an environment suitable for life.
f_l = the fraction of suitable planets on which life appears.
f_i = the fraction of life-bearing planets on which intelligent life emerges.
f_c = the fraction of civilizations that develop a technology that transmits detectable signs of their existence into space.
L = the length of time taken by such civilizations to release detectable signals into space.

Upon multiplying these seven variables, we obtain the number of planets in the Milky Way, which have possessed an intelligent civilization at some point in their history, rendering any predictions hypothetical for astrobiologists and extraterrestrial communicators alike. Nevertheless, the equation provides useful insights into what it takes to estimate how many civilizations exist in our galaxy [6].

Since 1961, the Drake equation's values have incorporated newly acquired scientific information. For example, when Drake wrote his equation, scientists did not know for sure whether stars other than the Sun had planets around them. Over the past few decades, advances in astronomy—and the discovery of exoplanets—have led scientists to understand how frequently a world with similar conditions to Earth might arise; we now know that there are at least 4000 exoplanets with the potential for habitability. With some ~200–400 billion stars in the Milky Way, we know that about one-fifth are Sun-like; about a quarter of those have Earth-sized planets orbiting with ~1-year periods. David Catling, an astrobiologist at the University of Washington, updated the Drake equation to estimate that there are at least four possible communicating civilizations in the Milky Way [6]. The estimate assumes that on such faraway planets, intelligent life arises in a way like it did on Earth. As Carl Sagan and Frank Drake once boldly proclaimed, 'civilizations more advanced than the Earth's exist elsewhere in the Universe' [12]. However, until we have discovered at least one example of extraterrestrial life, that conclusion cannot be considered secure.

19.5.1 SETI

Headquartered in Mountain View, California, the SETI Institute is a nonprofit research organization devoted to communicating with intelligent life in the Milky Way, if it should exist. The organization defines intelligent life as anything or anyone capable of building a radio transmitter sending signals detectable by civilizations in other worlds. SETI uses existing telescope arrays to determine whether anyone is sending out a message. It seeks to answer that question by hunting for advanced radio or optical signals in the vastness of the cosmos. Radio waves travel at the speed of light. At the right frequency, they pass entirely through interstellar space and penetrate a planet's atmosphere. If the largest radio telescope on Earth were pointing at an equivalent telescope on a planet around another star, then the two telescopes could be separated by thousands of light-years and still hear each other. So far, SETI has remained unanswered.

Until recently, the search for extraterrestrial intelligence has centered on detecting an incoming radio signal. With increasing computational power and more sensitive telescopes, researchers are targeting the technosignatures of advanced civilizations by expanding the search to optical and infrared emissions. These could include laser pulses, polluting gases, or megastructures built around a nearby star to harness energy. Although there are probably other civilizations in the galaxy with advanced radio telescopes, the vast number of stars and the vast distances of space remain daunting barriers to extraterrestrial communications. For more than half a century now, SETI has failed to detect any hint of alien civilizations. Nevertheless, it has gathered invaluable information in many branches of astronomy, including the detection and study of pulsars—small dense stars that can spin hundreds of times a second with enigmatic fast radio bursts.

Statistically considered, intelligent life so defined would be the rarest form of life. Still, the universe is vast. SETI is an activity of human exploration akin to Christopher Columbus' aggressive investigation of the Atlantic Ocean in the fifteenth century, leading to the discovery of the New World. The probability of detecting intelligent extraterrestrial life is difficult to estimate—but if we never search, the chance of success is zero. While the fascination with SETI waxes and wanes, our best and brightest astronomers persevere, searching for intelligent life. Private donors like Bernard Oliver, a prominent Hewlett Packard engineer, and Paul Allen, the cofounder of Microsoft, have supported SETI with generous funding. In 2015, the Breakthrough Prize Foundation, an organization created by the Internet investor Yuri Milner, signed a contract with the University of California, Berkeley, committing $100 million over the next 10 years to a Breakthrough Listen project designed to escalate the search for extraterrestrial life. The Breakthrough initiatives will radically accelerate the search for artificial signals from the nearest millions of stars and the trillions of stars farther out in the 100 closest galaxies.

Recently, SETI has been collaborating with other organizations, including America's National Radio Astronomy Observatory. The observatory operates a field of radio telescopes in New Mexico called the Very Large Array (VLA). The collaboration allows SETI researchers to tap into the entire stream of data that the VLA records as it carries out its routine observations. Similarly, SETI is also collaborating with the Square Kilometer Array (SKA), which has many dishes in South Africa and Australia. However, none have detected extraterrestrial signals so far. Intelligent life may just be too far away to hear us or for us to hear. Moreover, some ask, what about our planet makes us so unique that we merit the attention of intelligent extraterrestrials, assuming that they exist?

SETI has been searching the skies for techno-signals of extraterrestrial intelligence for nearly six decades now without a single peep, a situation that has been called 'the great silence.' There are several pieces to this puzzle. For one, the universe is so vast that SETI has searched only an incredibly small portion of it. Another is that SETI has been using twentieth and twenty-first century technology to contact much more sophisticated technical societies. Perhaps the aliens possess technologies considerably more advanced with respect to ours. Some argue that because the possibility of direct contact by interstellar travel is vanishingly little due to the enormous distances involved and the limitations imposed by the speed of light, communication with distant civilizations will demand radio communications that we must keep up for thousands of years. A signal from space would certainly have to be from a civilization that has existed much longer than ours. We are looking for a technology like our own, and so we presume that aliens will share our principles. However, an alien message may well be an attempt to communicate, but in a language that we do not understand. Perhaps, instead of communicating with us, aliens have sent probes to our solar system to observe or contact us. Some enthusiasts believe that aliens are already here, but we just do not know it yet.

'One day, we might receive a signal from a planet like this,' Stephen Hawking says in the documentary, referring to a potentially habitable alien world known as Gliese 832c. Hawking acknowledges that the discovery of intelligent life would be the greatest scientific discovery in human history. However, a visit from aliens may not end well, in the same way that Columbus' arrival to the Americas did not turn out well for the Native Americans. I agree with Hawking. I am highly interested in finding a signal that suggests the presence intelligent life elsewhere in the cosmos. However, I would be extremely wary of answering back.

19.6 Rare Earth?

Earth is the place where we know life arose, survived, prospered, diversified, and hosted techno-intelligence. Earth is a rare and precious place—perhaps the most fortunate planet in the cosmos. However, the abundance of life's building blocks in interstellar space suggests that many other planets in the solar system, the galaxy, and the universe may be hospitable and friendly to live. Throughout human history, people have believed that we are not alone. Recent decades have been proactive in the search for alien life and, thanks to a host of discoveries, that possibility now seems more plausible than ever. Any attempt to study extraterrestrial life's origin must be based on what we know about the origin and evolution of life here on Earth. We understand that the emergence of life primarily depended on the presence of water in

its liquid state. However, space exploration has shown how difficult it is to find liquid water in the solar system. The conditions required to maintain water in its liquid state are highly restrictive, and Earth is a privileged planet in this regard. However, given the superabundance of ice in the universe, liquid water should be common. If a planet or moon has internal heat and a hydrothermal system, ice can melt into water. NASA's exploration of Europa, Titan, and Enceladus gives us hope and encouragement as we explore the solar system and expand our quest for liquid water. Our search for life in the cosmos follows the water. In the distant past, water once flowed across the surface of Mars. New findings from NASA's Mars Reconnaissance Orbiter provide the most persuasive evidence that liquid water—albeit briny—still flows intermittently on Mars during warm seasons.

We assume that the building blocks of life are the same throughout the cosmos and that the processes operating on these building blocks are governed by the same physics and chemistry that rule Earth. If life's building blocks are widespread in the universe, then life should be too. Most astrobiologists believe that microbial life is widespread in the universe. Alien microbes may have the same biochemical basis as those that we know, but there is a large degree of contingency associated with the qualities of more complex organisms and human-style intelligence. Nevertheless, if we find evidence of life elsewhere—whether on Mars, Europa, Enceladus, or in the biosignature on a distant exoplanet—it will change everything, philosophically and epistemologically. The probability is strong that life also emerged elsewhere, wherever the necessary physical and chemical conditions were present. In this decade, we are discovering biomarkers of life on other planets and moons. Each year gives us more confidence that there is at least microbial life on many other habitable worlds beyond our solar system. De Duve believes that because life exists on Earth, it is a cosmic imperative that life must be universal throughout the universe [1]. Life has arisen under the conditions prevailing on Earth, and it will occur similarly whenever and wherever the same conditions prevail. De Duve agrees that chance played a role in biogenesis, but it was a chance tempered by various physical constraints that impose an overall directionality—with life as the predictable biological outcome. By this reasoning, some form of life likely exists on many of those countless exoplanets.

At the heart of the search for life elsewhere in the universe is the question: Is Earth unique? Are we the only technological species that has ever arisen on a given habitable planet? Is intelligent life a vanishingly rare event in the universe? Humans once believed that Earth is unique and lies at the center of the entire universe. Scientific investigations eventually showed that our world is not even at the center of its solar system—it is one of the eight planets and many smaller bodies orbiting the Sun. Our Sun is not at the center of the Milky Way galaxy but lies on a spiral arm 30,000 light-years from the center of the galaxy.

Are aliens real? Do aliens exist in the vast cosmos? In the summer of 1950, during a lunchtime conversation at the Los Alamos National Laboratory, the Nobel laureate Enrico Fermi, most famous for creating the first nuclear reactor, discussed the apparent contradiction between the lack of evidence for extraterrestrial civilizations and various high estimates of their probability (such as some of the more optimistic projections from the Drake equation). There are billions of stars in the Milky Way, like the Sun. With high probability, some of these stars have Earth-like planets. He pointed out that whereas Earth is only 4.5 billion years old, other stars in the Milky Way can be dated to 9 billion years ago. Thus, if intelligent life is common, then many technological civilizations should have arisen long before us. Earth should have been visited by intelligent aliens already. 'But where is everybody?' Fermi allegedly asked. This question is now dubbed the 'Fermi paradox': where are the aliens that should be out there? Although the Fermi paradox has baffled scientists for decades, some new insights help us understand why aliens are hard to find.

Like Fermi, many other scientists are skeptical about the existence of intelligent extraterrestrial life. For the Nobel laureate biochemist Jacque Monod, the author of *Chance and Necessity*, the genesis of life was overwhelmingly the product of chance, the result of a blind cosmic lottery: 'Man at last knows that he is alone in the unfeeling immensity of the Universe out of which he emerged only by chance' [13]. Life emerged on Earth quickly in its fertile post-Hadean window, but the arrival of a technological civilization took eons to accomplish. Drawing on our single data point—Earth—about four billion years elapsed until techno-intelligence developed on our planet. Life may be common, but radio technology is rare.

Our hominid lineage has an evolutionary history of seven million years since the emergence of *Sahelanthropus* in Africa. Our genus, *Homo*, is only 2.5 million years old. Our species, *Homo sapiens*, is just 300,000 years old. For most of our tenure on Earth, we were hunters and foragers wandering in small, tribal, extended families. Gradually abandoning hunting and foraging in the forest about 12,000 years ago, our ancestors in the Near East's Fertile Crescent, Far East's China, and India began agricultural cultivation and fundamentally changed the human lifestyle. Civilizations only arose about 10,000 years ago. Industrial revolutions, starting in 18th-century Great Britain, transformed our modern society. Our space age only began in the 1950s. In all our attempts to imagine intelligent life elsewhere, then, we must resist the anthropomorphic thinking that extraterrestrial intelligent life must be much like our own, or even better, be technology-wielding beings capable of galactic communication and

space exploration. There may be intelligent life elsewhere in vast space, but they may be intelligent in the manner of octopuses, dolphins, or whales, without technology. Dinosaurs reigned for 165 million years without any technological skill. Whether technological intelligence is unique to us or universally convergent is still impossible to know.

The question of evolution's predictability was notably raised by the late paleontologist Stephen Jay Gould, who considered evolution to be essentially contingent and unpredictable [14]. In *Wonderful Life*, Gould illustrates the thought experiment (*gedankenexperiment*) of replaying life's tape and seeing whether the outcome would be anything like the original. Gould's conclusion was, 'Replay the tape a million times…and I doubt anything like *Homo sapiens* would evolve again.' *Homo sapiens* resulted from a particular and unique combination of contingencies, a 'glorious accident' impossible to predict. Gould claimed that the history of life is characterized by contingency, the fact that every historical event depends on all the past events that led to it. Evolution is a robust historical process in which a change in any event along the way changes the cumulative outcome.

The biology on other planets will almost certainly differ from that of our own because of the statistical nature of the evolutionary process and the adaptability of life. It is true that the repeated evolution of similar traits, or functional similarities, is widespread in nature: insects, pterosaurs, birds, and bats independently invented wings for flight and fish, plesiosaurs, mosasaurs, and whales developed similarly streamlined bodies for swimming. However, even while homoplasy or the development of functional similarities can undoubtedly lead to different kinds of intelligent animals, it is still doubtful whether technological intelligence that has arisen in humans would appear in other lineages of life by convergent evolution.

So, is intelligent life a rare event in the universe, or, as many astronomers believe, are there many advanced civilizations throughout the universe? Eminent evolutionary biologists like George Gaylord Simpson and Ernst Mayr are skeptical about extraterrestrial intelligence, and even more so, if it exists, of any possibility of communicating with it. Paleontologist Peter Ward and astronomer Don Brownlee published a best-selling book on this matter, *Rare Earth*, stating that microbial life may be expected in the universe. Still, multicellular life may be so uncommon that Earth might harbor the only intelligent life in the Milky Way [15]. They called this idea the Rare Earth Hypothesis: Earth-like planets genuinely suitable for intelligent life are few and far between. Microbes do have a remarkable range of metabolisms and can survive in a far wider range of environments than more complex organisms, and newly discovered exoplanets are potential refuges for alien life. Billions of environments surely exist in our galaxy where life is possible, but technologically intelligent life would still be a rare event.

19.7 Conclusions

Humanity has long wondered whether we are alone in the universe. As NASA has explored our solar system and beyond, it has developed increasingly sophisticated tools to address this fundamental question. There is a strong possibility that microbial life also originated beyond Earth whenever the necessary physical and chemical conditions were met. Understanding life on Earth should help us find life on other planets and moons. Life on Earth depends on water. To find life elsewhere, we have to look for water. Water in its various forms permeates the solar system. Jupiter's moon Europa is covered with a sheet of ice, and Saturn's moon Enceladus shows evidence of subsurface water. Mars was once a relatively warm and wet world in the Noachian that apparently harbored large amounts of liquid water, the remnants of which are now frozen. The discovery that extremophiles can survive under exceedingly harsh conditions has broadened scientists' anticipation of where life beyond Earth might reside. The conditions for life are plausible on a handful of other habitable places in our solar system, such as Mars, Jupiter's moon Europa, and Saturn's moons Titan and Enceladus.

However, while the hunt for life on nearby planetary bodies is thriving, our solar system is not the limit of these efforts. The Kepler space telescope has recently discovered about 5000 Earth-like exoplanets in the galaxy, each orbiting in the habitable zone around a Sun-like star. The systematic search for extraterrestrial intelligence (SETI) in our galaxy has proceeded for over half a century. Radio telescopes are searching for signals produced by intelligent life. At the same time, universal contingency has been cited as an important argument for our uniqueness as an intelligent species and our planet's uniqueness in supporting intelligence. While we probably will not find intelligent life in our galaxy anytime soon, there is a good chance that we may see microbial life in our solar system. The question is no longer, is there life beyond Earth, but, rather, how do we find it?

References

1. De Duve C (2005) Singularities: landmarks on the pathways of life. Cambridge University Press, New York
2. Greaves JS, Richards AMS, Bains W, Rimmer PB, Sagawa H, Clements DL, Seager S, Petkowski JJ, Sousa-Silva C, Ranjan S, Drabek-Maunder E, Fraser H, Cartwright A, Mueller-Wodarg I, Zhan Z, Friberg P, Coulson I, Lee E, Hoge J (2020) Phosphine gas in the cloud decks of Venus. Nature Astron. https://doi.org/10.1038/s41550-020-1174-4
3. Carr MH, Head JWH (2010) Geologic history of Mars. Earth Planet Sci Lett 293:185–203
4. Grotzinger JP et al (2014) A habitable fluvio-lacustrine environment at Yellowknife Bay, Gale Crater, Mars. Science 343. https://doi.org/10.1126/science1242777

5. Kate ILT (2018) Organic molecules on Mars. Science 360:1068–1069
6. Catling DC (2013) Astrobiology: a very short history of introduction. Oxford University Press, Oxford
7. McKay DS, Gibson EK, Thomas-Keptra KL, Vali H, Romanek CS, Clernett SJ, Chiller XDF, Maechling CR, Zare RN (1996) Search for past life on Mars: possible relic biogenic activity in Martian meteorite ALH84001. Science 273:924–930
8. Thomas-Keprta KL, Clemett SJ, McKay DS, Gibson EK, Wentworth SJ (2009) Origins of magnetite nanocrystals in Martian meteorite ALH84001. Geochem Cosmochem Acta 73:6631–6677
9. Waite JH, Glein CR, Perryman RS, Teolis BD, Magee BA, Miller G, Grimes J, Perry ME, Miller KE, Bouquet A, Lunine JI, Brockwell T, Bolton SJ (2017) Cassini finds molecular hydrogen in the Enceladus plume: evidence for hydrothermal processes. Science 356:155–159
10. De Sanctis MC, Ammantio E, McSween HV, Raponi A, Marchi S, Capaccioni F, Capria MT, Carrozzo FG, Clarniello M, Fonte S, Formisano M, Frigeri A, Giardino M, Longobardo A, Magni G, McFadden LA, Palomba E, Pieters, Tosi F, Zambon F, Raymond CA, Russel CT (2017) Localized aliphatic organic material on the surface of Ceres. Science 355:719–722
11. Shklovskii IS, Sagan C (1966) Intelligent life in the universe. Holden-Day, San Francisco
12. Sagan C, Drake F (1997) The search of extraterrestrial intelligence. Sci Am 276(1)
13. Monod J (1970) Chance and necessity: an essay on the natural philosophy of modern biology. Knopf, New York
14. Gould SJ (1989) Wonderful life. Norton, New York
15. Ward P, Brownlee D (2000) Rare Earth: why complex life is uncommon in the universe. Copernicus, New York

Epilogue

20

Life arose through a long succession of chemical steps that were bound to take place under the conditions that prevailed on Earth four billion years ago.

—Christian de Duve, 1995.

20.1 The Longest Voyage

Life emerged four billion years ago, but its origin is lost in the mists of ancient eons. We have tried to reconstruct the biochemical pathways for the origin of life on young Earth through multiple lines of evidence and conjecture. We have presented a broad outline of how life on Earth might have begun. Meteorites that bombarded our planet brought organic compounds, concentrated and combined in suitable hydrothermal cradles, to make cells capable of reproducing and becoming the increasingly complex life forms we see today. However, the way in which the first cells evolved still remains unknown. We know that living cells' intrinsic properties do not come from individual cell components such as nucleic acids, ribosomes, proteins, or cell membranes but from their collaborative dynamics mediated by sophisticated information systems. Self-organized lipid structures (protocells) have been proposed as an intermediate stage between nonliving materials and cellular life. The synthetic production of model protocells can demonstrate the potential processes by which living cells first arose. Suppose that we succeed in creating artificial cells like LUCA, the last universal common ancestor. In that case, we might then understand how these primitive cells formed and started to divide and evolve.

Like any good story of a voyage, our narrative of life's origin has a beginning, a middle, and an end. To quote Arthur C. Clarke, the renowned science fiction writer and futurist. 'Across the gulf of centuries, the blind smile of Homer is turned upon our age.' We have now come to the end of this voyage in the search for life's ultimate origin. In the beginning, we had a cosmic connection. The solar system formed 4.55 billion years ago out of a collapsing cloud of dust and gas left over from a supernova explosion that already had a rich composition of organic material, as preserved in the carbonaceous chondrites, comets, and interstellar dust that delivered the building blocks of life to Earth. We began our voyage with complex cosmic biomolecules that could have seeded the ingredients of life on Earth. These cosmic, organic molecules had a built-in analog system.

The middle period of our quest is murky and has left few records. We have reconstructed it from astrobiological records and experimental evidence. The central part traces the chemical evolution of protocells via the analog, hybrid, and digital information stages, accelerated by the interaction of ancient virus world. Despite knowing approximately when the first life appeared on Earth, we are still far from answering how it emerged from the prebiotic world. A plausible account of these events is the long and complicated rising action of the narrative. The third part traces the origin of the first cells from DNA protocells. However, by the end of the voyage, the most monumental event in the history of our planet occurred, i.e., the emergence of the first cells, thus making Earth a living world.

In this book, I have presented a bottom-up, integrative approach to retrace the steps to the origin of life from stardust to the emergence of the first cells (Figs. 4.1 and 17.2). We have gone through five successive ages corresponding to five levels of complexity: the cosmic age, the geological age, the chemical age, the digital information age, and the biological age. These ages overlap as one gradually leads to the next. In the cosmic age, the building blocks of life had their beginnings in interstellar space during a supernova explosion. During the Late Heavy Bombardment period, both comets and carbonaceous chondrites delivered the building blocks of life (with an embedded analog information system) and water to early Earth. The age of geology discusses the likely cradle, such as hydrothermal impact crater lakes, for the beginning of prebiotic synthesis. The age of chemistry concerns a long evolution of protocells and biopolymers and contains the essence of life as a chemical process. Noncoding RNAs gave rise to the hybrid information system. The age of digital information in the peptide/RNA world gives us the

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5_20

algorithms and logic of life, first with the genetic code and translation system.

The advent of proteins saw the return of analog information along with the development of the plasma membrane, protein folding, and enzyme–substrate interaction. Protocells with plasma membranes acquired remarkable abilities to process analog information to sense, respond, and adapt to their internal and external environments. The availability of protein leads to a virus world, a significant detour. A virus itself, in its dormant state, is a hybrid of RNA with a capsid coat but gets activated and self-replicates during infection of protocells. Ancient viruses invented horizontal gene transfer (HGT) and horizontal enzyme transfer (HET) to other protocells. Without the assistance of viruses, abiogenesis might have taken a different trajectory without the emergence of DNA. Viruses are the unsung but dynamic players in the evolution of life. With the advent of viruses, we saw the first evidence of the transfer of digital information from a virus to a protocell. Viruses moved RNA, DNA, and all critical enzymes between protocells, provided new genetic materials for abiogenesis, and regulated vast populations of early cells. They transferred DNA and all critical replicating enzymes to protocells during recurrent infection [1]. The coevolution of protocells and primitive retroviruses in the protein/RNA world is supported by an experimental study. Bansho et al. [2] synthesized a simple replication system composed of two self-replicating host and parasitic RNA species. When the system was compartmentalized, a continuous oscillation pattern in the two RNAs' population dynamics emerged. The authors concluded that the origin of host–parasite coevolution might have appeared in the very early stages of the origin of life.

DNA transcription in protocells enlarges the scope of digital information with the establishment of the central dogma, where the flow of information passes from DNA to mRNA to proteins. DNA replication mediates the first cell division with an intricate choreography in which analog, hybrid, and digital information passes from parent to daughter cells accurately. The prebiotic journey ends with the emergence of the first life with the intricate dance of cell division. The signature of each life is cryptically encoded in the nucleotide sequences of DNA. Life learns how to propagate information from generation to generation and thus how to evolve.

Other than nucleotide sequences, DNA has a second type of digital information, i.e., the gene regulatory network (GRN), which is a set of genes that interact with each other to control a specific function (Fig. 17.1). It governs the gene expression level of mRNAs and proteins, which in turn, determine the function of the cell in response to environmental changes, such as nutritional status and several stresses.

Life comprehends the interaction of all the inanimate components of this whole system. In other words, nothing is alive in a cell, except the whole system itself. Crucial cell components, such as the plasma membrane, nucleic acids, and proteins, though sophisticated and organized, are not alive, and they obey the same physical and chemical laws as do inanimate systems. Life is an emergent property whose systemic behavior cannot be understood or predicted from its individual parts or cell components alone. Emergent properties necessarily arise from the collective interactions of the parts within a more extensive system. For instance, living systems are composed of biopolymers, such as nucleic acids and proteins, which interact within a lipid bilayer membrane with three primary functions: containment, transport of nutrients, and energy transduction. Nucleic acids such as mRNA and DNA have a unique ability to store and transmit genetic information. Similarly, protein enzymes have a unique ability to catalyze metabolic reaction rates. Genetic and catalytic polymers joined in a symbiotic relationship that increased the efficiency of both biopolymers. Life began when a few of the immense numbers of protocells incorporated a symbiotic cycle involving genetic information and catalytic functions.

There may be multiple pathways to the origin of life, but we know the beginning and end of this long voyage and the major steps in abiogenesis that led to the emergence of the first cells. Following the hierarchy of life synthesis, our next task is to look for any underlying pattern and rule that might have given rise to the first life from the cosmic molecules delivered to young Earth about four billion years ago.

20.2 The Chaotic Origin of Life

Science has long been based on the concept that the order of the universe, from the infinitely small to the infinitely great, from atoms to galaxies, is genuinely universal. Such grand simplicity comprehends the complexity of the details of a living organization. The relative orderliness of our cells does not reflect any elegant design. How does the organized form of a living cell emerge from the chaotic motion of its constituent components and chemicals? Is the origin of life a cosmic chance or a cosmic necessity? Is the pathway that goes from nonliving to the first cells due to a unique event resulting from chance operating in a particular time/space situation? Or is it fated to occur by the laws of physics and chemistry? This dichotomy between chance and determinism is a profound and long-debated question in the philosophy of science, with good arguments on both sides.

For instance, Jacques Monod sees life's origins in general and our existence as a matter of chance, unlikely to be replicated anywhere in the universe [3], a view of contingency echoed by other scientists [4, 5]. Even if microbial life anywhere in the cosmos may have the same chemical basis, a large degree of further contingency must still be associated

with multicellular organisms and technological intelligence [6]. However, most astronomers and planetary scientists believe that microbial life is widespread in the universe. According to Christian De Duve, life's origin is a cosmic imperative, an inevitable and deterministic event, and an obligatory manifestation of cosmic chemistry bound to arise elsewhere in the universe where conditions are appropriate. Given the suitable initial conditions, life's emergence is highly probable and governed by the laws of physics and chemistry. Moreover, the constraints of physics and chemistry are so strong that life in any other place would be the same [7].

In a similar vein, paleontologist Stephen Jay Gould claims that life on Earth evolved quickly under the right conditions. The first microbial fossils appeared soon after Earth had cooled down sufficiently. Gould believes that given early Earth environment's chemical composition and the physical principles that govern self-organization, life's origin was virtually inevitable [8]. Morowitz suggests that biogenesis was a series of chemical events subjected to the regular laws governing atoms and interactions [9]. He further emphasizes that by assuming deterministic processes were at work on the planet, we can pursue the understanding of life's origins within the constraints of science.

In my view, the moment of the origin of life is to be considered a singular, nonlinear, chaotic event of historical contingency, an outcome of long chains of unpredictable antecedent states at the intersection of chaos and order. Chaos is inherently uncertain, random, without form, potential. Order is predictable, structured, actual. First life was highly ordered, but abiogenesis was largely chaotic but approaching toward order. Before life emerged, many physicochemical processes on our planet were highly chaotic. Life brings order out of chaos. Subsequent life on Earth is always unique, changing, and unpredictable. However, out of these chaotic assemblages, directionality emerged during the prebiotic synthesis, as suggested by De Duve and Morowitz. In the next section, we discuss directionality in the history of life at different levels, which might have accelerated the origin of life synthesis. Chaos theory may provide a new understanding of how complex patterns arise during the emergence of life. In this theorization, chaos is a deterministic phenomenon sensitive to initial conditions leading to an unpredictable/aperiodic behavior [10]. Determined by a complex interplay of myriads of biomolecules, all influencing one another's behavior and environments, life's origin has many characteristics typical of such nonlinear, chaotic systems.

First, a chaotic system is based on sensitive dependence on initial conditions (SDIC). Extraterrestrial delivery of organic compounds in conjunction with geochemistry must have defined the types of molecules present on early Earth, the molecular composition of early chemical systems, and, by extension, that of protocells and contemporary cells. Furthermore, the environmental conditions must have defined the potential reactivity of these compounds. The starting point was the availability of specific cosmic ingredients and liquid water on young Earth, fortunately within the habitable zone of its Sun. Second, prebiotic synthesis is chaotic in the cosmic and geological stages, then deterministic in the chemical stage, but becomes more stochastic but directed in the digital information and biological stages. For example, changes in one part of the system, such as the nucleic acid base sequence, directly caused other changes, such as the composition of amino acids and the structure of protein molecules. Other attributes and processes of chaotic systems, such as self-similarity and feedback, are both critical features of life. However, the outcomes of such changes cannot be known beforehand. Just like the weather, changes are inexorable but can only be followed with the benefit of hindsight.

Third, during the emergence of life from extremely simple organic compounds to the first cells, there was an increase in the complexity of biomolecules. During the prebiotic process, natural selection rejected many molecules not suitable for biogenesis, recycling the unused ones to develop new raw material. In the digital information age, the amount of new information perpetually increased during life synthesis through self-amplifying feedback loops with the creation of the genetic code, which further pushed the system away from equilibrium until it reached a threshold of stability, a bifurcation point, that is a point of instability. At this point, new forms of order may emerge spontaneously, resulting in development and evolution. The advent of proteins was such a watershed event in biogenesis. It created two parallel but interconnected worlds—a virus world and a protocell world—leading to the DNA world by repeated viral infections. Finally, in the first cells capable of reproduction and Darwinian evolution (Fig. 20.1), reproduction was not perfect, so new genetic variations arose by mutation—the raw material for natural selection—resulting in differences in a population of cells.

Fourth, the singular pathway for the emergence of life was directed and constrained by an unpredictable historical contingency. As with the weather, a chaotic system continually generates new information unavailable at the start of the process. Although each change on such an evolutionary path has some causal relation to the circumstances in which it arose, outcomes must eventually depend on the details of long chains of initial conditions or antecedent states and small changes that may have enormous long-term repercussions. Evolution and life itself incorporate all these qualities of chaotic systems.

At its core, the origin of life involves a profound tension between random and deterministic processes. From dying stars to the origin of life, each change on the evolutionary path has some causal relationship with the circumstances in which

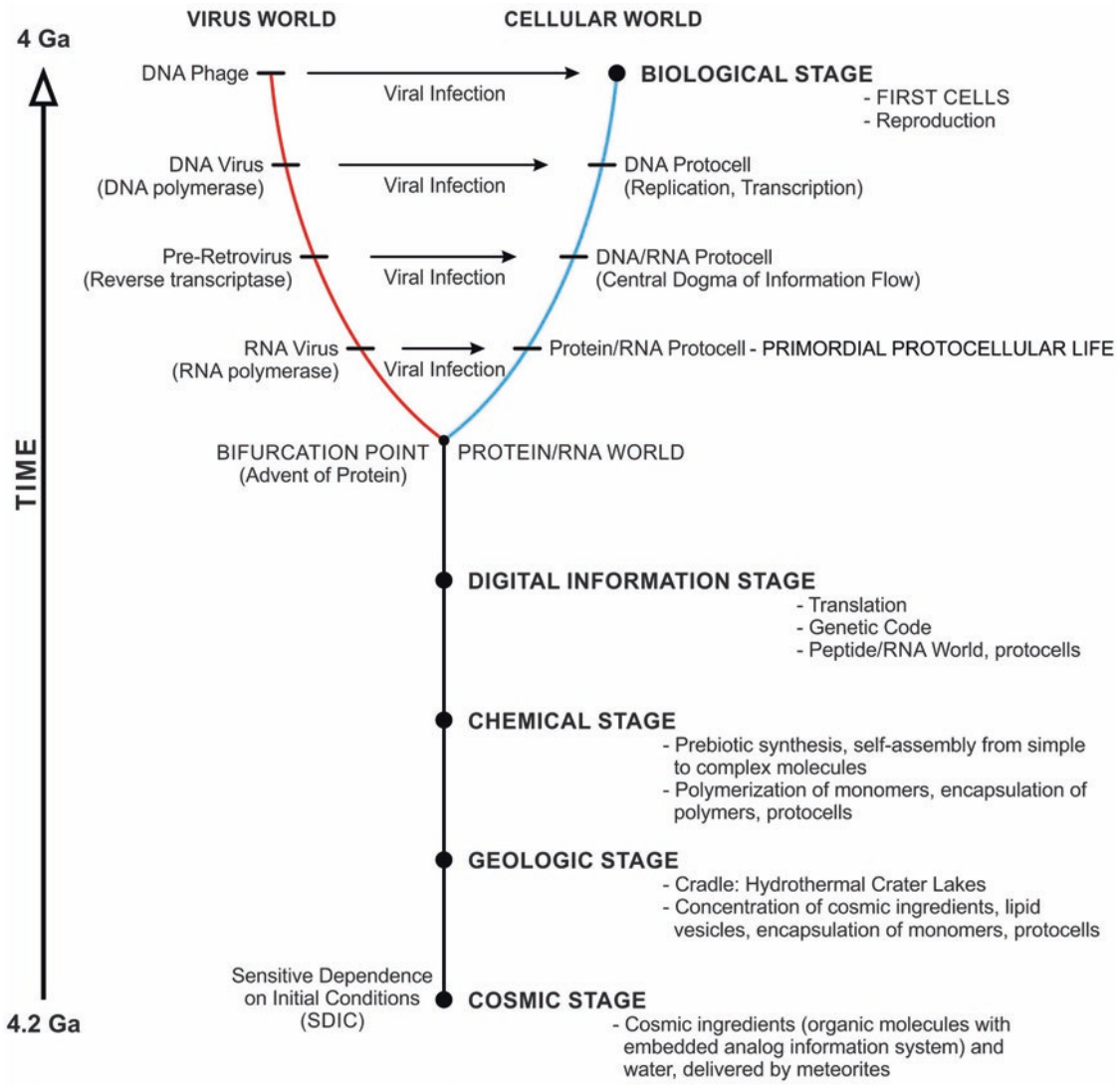


Fig. 20.1 The origin of life shares many typical characteristics with a nonlinear chaotic system. It started with the cosmic stage when the cosmic ingredients (with embedded analog information) and water molecules were delivered to young Earth by asteroids during the Late Heavy Bombardment period. This cosmic stage represents the sensitive dependence of initial conditions (SDIC), a key component in any chaotic system. Even in an entirely deterministic system, a slightest change in the initial data can cause abrupt and seemingly random variations in the outcome. The Miller-type experiments failed to produce more complex biomolecular networks because their SDIC was incorrect. The following three stages in succession represent the self-assembly of complex molecules from simple monomers and the subsequent genetic code and translation apparatus. Digital information systems started in the peptide/RNA world with the emergence of the translation system and the genetic code. Another milestone in prebiotic synthesis was when proteins and various enzymes were produced during the decoding and translation of mRNAs. The availability of a wide range of proteins created a distinct bifurcation point, creating two parallel abiogenesis pathways, namely, the virus world and the cellular world, which coexisted and intermixed. Viruses began to infect protocells and enriched their genomes and immune system. Donation of an RNA replicase such as RdRp by mRNA virus to protein/mRNA protocell led to the primordial protocellular life. In the retro world, pre-retrovirus, while infecting protocells, donated DNA and critical enzymes to protocells for replication of RNA and DNA. The arrows between the two worlds indicate the continuous horizontal transfer of genes and enzymes between cells and viruses by repeated infections. The climax of the chaotic system of early abiogenesis was the emergence of the highly ordered first cells capable of reproducing identical daughter cells

it arose. Natural selection systematically works to adapt populations of biomolecules to their prevailing environments, but the emergence of complexity in biomolecules gives directionality to their evolution. In many molecular settings, the range of possible pathways may be limited, yielding what appears to be reasonably deterministic processes. In biogenesis, the first three stages—cosmic, geological, and chemical—were somewhat random, depending on the environment and the biomolecules' availability. The basic building blocks of life could be assembled anywhere in interstellar space. However, for life to

emerge, these building blocks require conditions similar to those on Archean Earth. When information entered the scene, the prebiotic chemical system became more deterministic. However, then, the process became random again when the first cells began to divide, mutate, and undergo Darwinian evolution. Moreover, although mutations, the first drivers of evolution, occur at random, nature often finds the same solution to the same problem. The early microbes might respond to environmental changes in identical ways. A recent work has suggested that there is an element of predictability in the randomness of evolution [11].

Thus, both random and deterministic processes became intertwined in abiogenesis such that future alternatives became contingent on the previous history of the evolving population. The roles of contingency and determinism of Earth's biogenesis should be relevant to understanding life trajectories elsewhere in the universe. Hidden in this nonlinear, chaotic life system is an upsurge of orders through self-organization. From cosmic biomolecules to the first cells, life comes about as a chaotic system that is unpredictably predictable for a while and then orderly in its randomness. Life itself is an example of emergence property. For instance, the first cell, like a single-celled bacterium, is alive but if we separate these macromolecules that combined to create the bacterium, these units are not alive.

20.3 Coevolution of Biomolecules and Prebiotic Information Systems

The secret of life lies in the exotic chemistry of cosmic ingredients and the emergence of information-based directionality and complexity that created order out of chaos. The living cell is an information-processing system that uses sophisticated computational code. Information is the key property of life, but the role of information as it operates in abiogenesis is less appreciated. Informational processes in life emerged from the environment and were elaborated in the prebiotic milieu of chemical networks. Information travels unidirectionally through biomolecules. Biomolecules whisper with each other – collecting inputs and generating outputs from the information systems. This is why life has evolved computation, language, communication, organization, memory, architecture, autonomy, self-recognition, and reproduction. We have identified two cycles of coevolution of prebiotic information systems with biomolecules. The first cycle shows the prebiotic synthesis from cosmic ingredients to the mRNA-directed proteins (Fig. 3.1). The second cycle shows the abiogenesis from proteins to the first cells via recurrent viral infection of protein/RNA protocells (Fig. 20.2). We have discussed how different kinds of biomolecules are embedded in different kinds of information systems such as analog, hybrid, and digital.

Although the analog information system (AIS) and the digital information system (DIS) are well-known in the literature [12], we have identified a transitional information form, the hybrid information system (HIS), composed of noncoding RNAs [13]. The coevolution of biomolecules and the AIS created noncoding RNAs by polymerizing nucleotides that gave rise to the hybrid information system (HIS). The HIS employed different species of noncoding RNAs such as ribozymes, pre-tRNA and tRNA, ribosomes, and functional enzymes, including bridge peptides, pre-aaRS, and aaRS—the different components of the translation machine.

The HIS gave rise to a digital information system (DIS) where aa-tRNA became the molecular architect for creating the first gene in two-step processes. First, they created codons of mRNA by base pair interaction (anticodon–codon mapping). Second, each aa-tRNA transferred its amino acid information to the corresponding codon (codon–amino acid mapping). aa-tRNA was the 'first molecular transformer' that created the first gene step by step. With the advent of encoded mRNA molecules, the first gene emerged before DNA. Molecular memory would play a critical role in assembling the first gene [14]. With the genetic memory permanently embedded in the digital sequences of mRNA, a mapping mechanism was developed between each codon and its cognate amino acid. As more and more codons 'remembered' their respective amino acids, this codon–amino acid mapping system developed the genetic code in their memory bank (Fig. 3.1).

The climax of the digital information system was the translation of linear digital sequences of mRNA into a three-dimensional protein in the analog format. aa-tRNA, in collaboration with the ribosome, was the 'transformer' that translated the mRNA language sequences into protein language sequences. It was a watershed event in the origin of life. Three prebiotic information systems, the DIS, HIS, and AIS, worked in harmony for this remarkable achievement of making proteins (Fig. 3.1).

Digital information of mRNA paved the way for the second wave of analog information by manufacturing custom-made proteins according to the specificity of the codon sequences of mRNA Two significant events followed with the availability of proteins: the evolution of phospholipids and plasma membrane of protocells and their protein-rich cytoplasm. The enzymes that evolved during this mRNA-directed process considerably enhanced chemical reactions and accelerated abiogenesis (Fig. 20.2).

The advent of proteins heralded the ancient virus world when viral genes began to concentrate in hydrothermal vent environments and were randomly coated by proteins for durability and protection. A virus is a hybrid molecule of nucleic acid and protein capsid. This was the beginning of mRNA viruses. Once the capsid protein began to coat the viral gene, the first mRNA virus appeared. In the hybrid state, a virus is

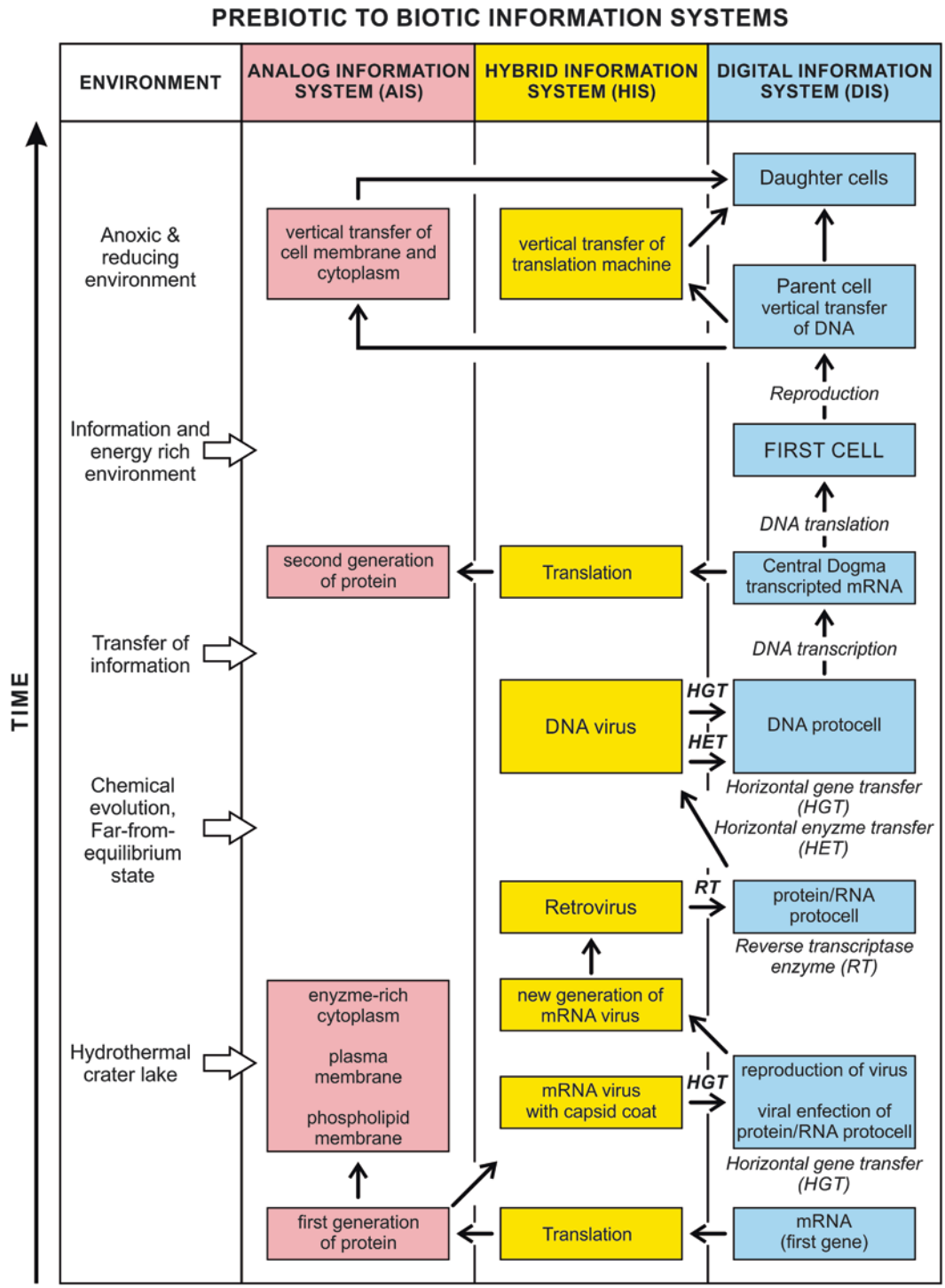


Fig. 20.2 Coevolution of biological molecules with biological information system in the protein/RNA world. An analog system with newly generated proteins began to coat mRNA molecules in the hydrothermal crater vent environment for stability and durability. This hybrid component was the early manifestation of a primitive virus and remained dormant. As these ancestral virus particles began to infect protocells, they hijacked translation machinery to produce viral proteins and replicated within protocells to create a new generation of mRNA viruses. Viruses provided a critical RNA-dependent RNA replicase (RdRp) enzyme during infection of protein/RNA protocell that facilitated the replication of *mRNA* gene and created primordial life. Virus also played critical roles in creating DNA protocells by horizontal transfer of genes (DNA) and donating critical enzymes for transcription and replication. The one-way flow of information from DNA to mRNA to proteins is called the central dogma of molecular biology. DNA replication, on the other hand, would enhance the cell division process. Cell division of the first cell exhibits how digital, hybrid, and analog information systems are passed vertically from parent to daughter cells

inert and nonliving. It gets activated and self-replicates during protocell infection by horizontal gene transfer (HGT). Gene exchanges between viruses and hosts are considered the major drivers of evolution. These primitive viruses began to infect protocells and exploited their ribosomes to create viral proteins. They also provided RNA-dependent RNA replicase (RdRp) to protein/mRNA protocell by horizontal enzyme transfer (HET) and facilitated the symmetrical division of a protocell to daughter cells. This was the beginning of the primordial protocellular life when viruses became 'the second transformer' in the origin of life (Fig. 20.2).

The next stage in viral evolution was the emergence of a primitive retrovirus with a new kind of replicative strategy in the sense that it could turn its RNA into DNA using its own reverse transcriptase (RT) enzyme that facilitated the transition from RNA to DNA genomes. With persistent infection, DNA viruses are slowly transferred from protein/RNA protocells to protein/DNA protocells by HGT. DNA viruses also transferred their core replication enzymes by horizontal enzyme transfer (HET). Thus, began the 'DNA world' when DNA replaced RNA as the major genome of the protocells. This is the continued role of the virus as the 'second transformer' in the origin of life.

DNA transcription in protocells enlarges the scope of digital information with the establishment of the central dogma, where the information flows from DNA to RNA to proteins. DNA replication mediates the first cell division with an intricate choreography in which digital, hybrid, and analog information passes from parent to daughter cells accurately. With the emergence of the first cell, DNA-based cellular life replaced the protein/mRNA based protocellular life by competitive exclusion (Fig. 20.2).

Two molecular 'transformers' played critical roles in creating nucleic acids (mRNA and DNA), heralding two distinct digital ages, enhancing abiogenesis, and providing directionality to achieve order out of chaos. The first 'transformer' was aa-tRNA that created mRNA or the first gene step-by-step. This was the first digital age, when a protein chain was synthesized from the encoded mRNA by translation, giving rise to the universal genetic code (see Chap. 12). The second 'transformer' was retrovirus that converted protein/RNA protocell into DNA protocell by reverse transcriptase enzyme, thus heralding the second digital age. Retrovirus donated not only DNA genes but also critical enzymes to protein/RNA protocell during upgrading of mRNA to DNA. In this second stage, mRNA is transcribed directly from a segment of DNA to build a protein; aa-tRNA was no longer required for building mRNA. DNA replication orchestrated identical cell divisions that led to the emergence of the first life. Once DNA appeared in the scene, the role of retrovirus and its critical enzymes were no longer necessary (see Chap. 16). These enzymes were created by early cells. Here we see the paradigm of information systems during abiogenesis—preserving necessary information to the next level and destroying enormous streams of unnecessary complex information, such as the primary functions of two transformers. These transformers were detours in abiogenesis, and they were abandoned when the goal was achieved by direct route. This reduction of information is purposeful, useful, and necessary for the survival of the first cell (see Sect. 3.4). This is why the origin of life study is too difficult to comprehend because of a large amount of missing data and detours, which were systematically erased hierarchically when the next level was reached, leaving no clue in the biochemistry or molecular biology of recent bacteria. Moreover, it had destroyed many biomolecules and their embedded information systems, when a goal of a certain level of abiogenesis was reached. The origin of life study has revealed how natural selection might have worked even at the molecular level.

We believe that the tripartite division of biological information systems into the AIS, HIS, and DIS represents a significant change of focus on the origin of life. These three information systems, operating separately and in close cooperation, streamlined the prebiotic synthesis from chaotic molecular assemblages and provided directionality to the flow of information. Information is an integral part of life. When we see the beauty and complexity of life, information systems might be the hidden cause that provided directionality, organization, and structure.

20.4 Directionality in the Hierarchical Origin and Early Evolution of Life

The way that information flows through biomolecules has led to accelerating the chemical evolution and has provided directionality and complexity in abiogenesis and the emergence of the first life. The origin of life has produced organic molecules of increasing size and complexity through time. How can a living system emerge from a chaotic assemblage of space molecules in a hydrothermal vent environment? What rules might have guided the prebiotic synthesis? De Duve [7] viewed pathways of life as both determinate and directional, where the vector of evolution lies in the structural, informational, and catalytic molecules (Fig. 20.3).

20.4.1 Directionality in Digital Information Systems

The digital information system and its unidirectional information flow in the prebiotic system are well known (Fig. 20.1). Some significant steps of digital information, as

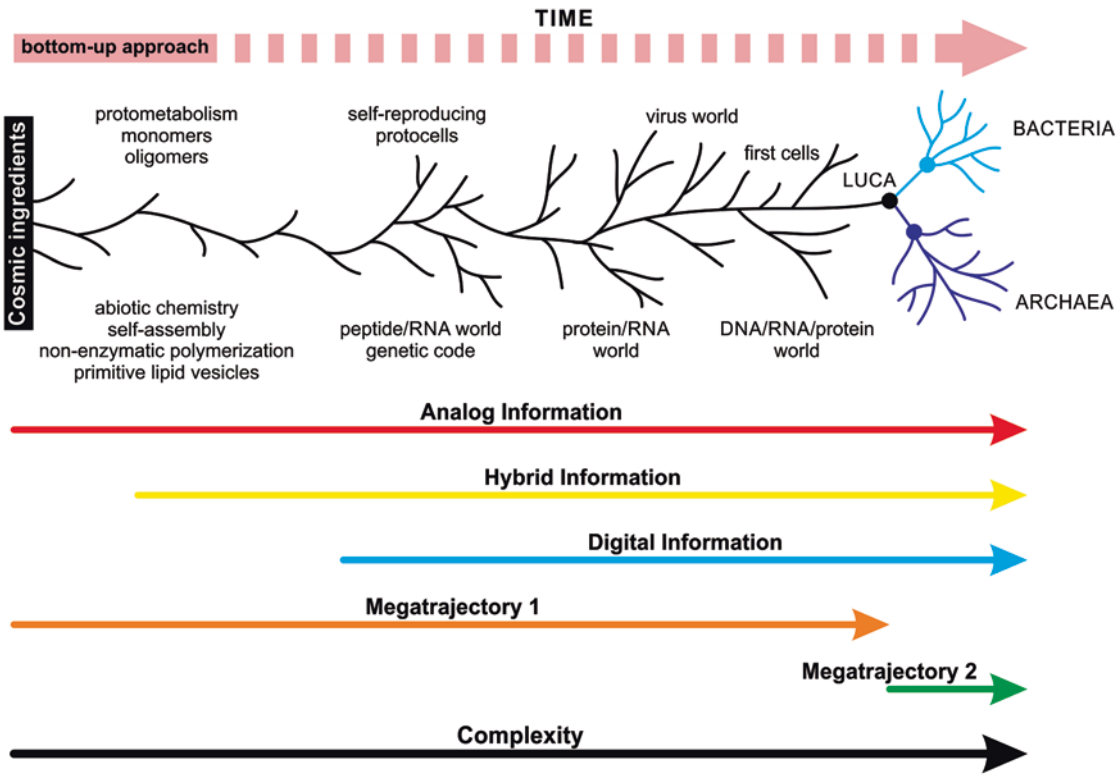


Fig. 20.3 The significant steps in the hierarchical origin of life using the bottom-up approach from stardust to LUCA show the directionality in life's history, driven by the analog and digital information systems; from LUCA, two domains of Archean life—bacteria and archaea—branched out. The evolution of life starting from the origin of life to LUCA represents the first megatrajectory, whereas LUCA to the metabolic diversification of bacteria and archaea represents the second megatrajectory [15]

it becomes organized in each level of hierarchical evolution, are as follows: (1) in the peptide/RNA world, with the origin of translation and the genetic code and information flow from mRNA → proteins; (2) in the retro world with information flow from mRNA → DNA → mRNA → proteins, (3) in the DNA world with information flow from DNA → mRNA → proteins, and (4) in the reproduction of the first cell. All three information systems, analog, hybrid, and digital, are replicated and transferred vertically to each daughter cell unidirectionally: parent cell → daughter cells. Both viral infection of bacteria and their immune system are digital in nature (Fig. 20.4).

Life is an open system, and energy such as adenosine triphosphate (ATP) must be fed at each step to improve the flow of digital information. Layer after layer of transitional digital information flow ratcheted the system up the complexity ladder toward the emergence of the first life. When the stable memory of digital information was permanently embedded in the first cell, a universal information system in all life emerged. Similarly, as efficient reproductive mechanisms are established in LUCA, it is split into two domains of bacteria and archaea by HGT with distinctive genetic and metabolic pathways (Fig. 17.6).

Directionality is also embedded at a lower molecular level in the transcription and translation of DNA. In transcription, the RNA polymerase synthesizes the mRNA strand in the 5′ → 3′ direction; similarly, in translation, the ribosome moves along the mRNA in the 5′ → 3′ direction. Thus, digital directionality occurs at each hierarchical level from nucleic acids to LUCA.

Digital information always produces analog output. In many cases, life's information system is hybrid, an inextricable mixture of digital and analog components. Digital information cannot operate without the assistance of analog replicating enzymes. Living systems are open systems with an unlimited energy supply from the environment. Energy from ATP, produced by analog information by chemiosmosis, is used to fuel chemical reactions, including those required for transcription of DNA to mRNA and translation of mRNA to make proteins. However, ATP is a nucleotide, one of the building blocks of RNA. Similarly, the analog cue from the environment prompts protocells to create proteins or divide.

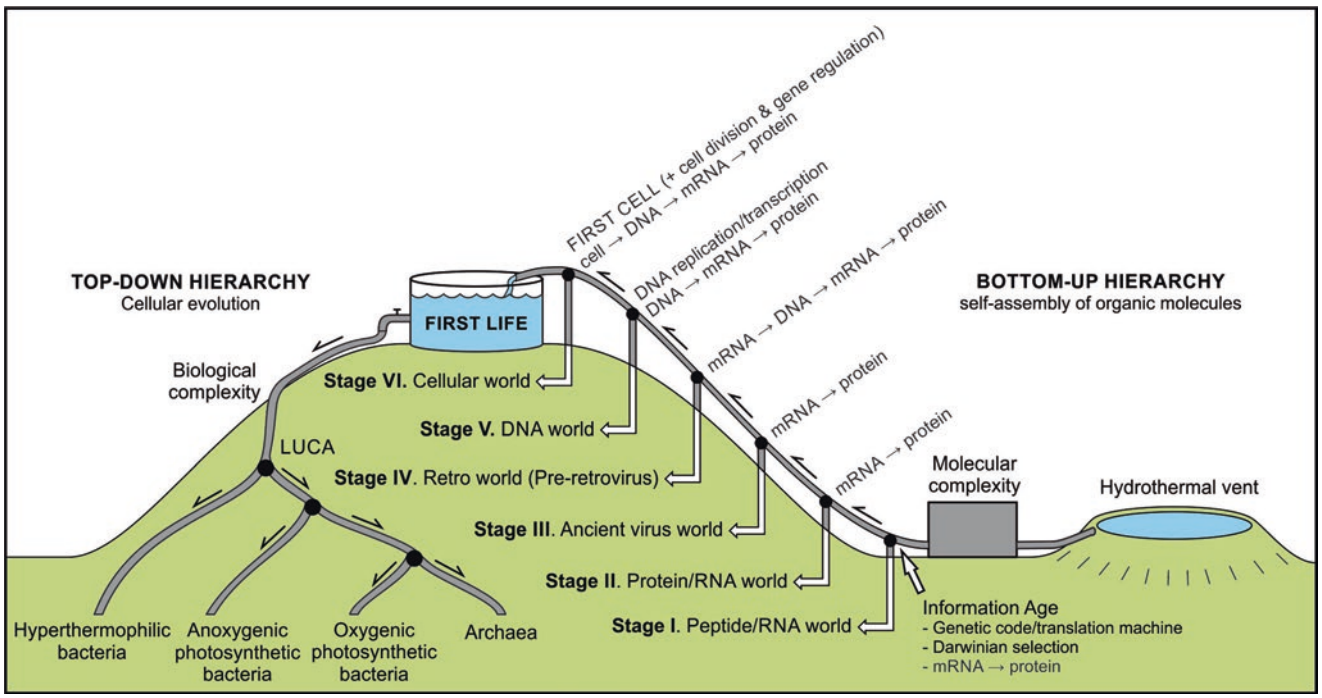


Fig. 20.4 The coevolution of life molecules and their digital information system emerged in an uphill battle in several key steps during abiogenesis. The piecemeal buildup of digital information flowing uphill during abiogenesis was fueled by analog ATP created in the hydrothermal crater vent environment. When the first life emerged, the accumulated digital information became permanently embedded in nucleic acids and flowed downhill in all life. Surprisingly, the major contributions of viruses in the final assembly of the first life and its information system are completely obliterated in modern cellular life, except for remnants of a few critical enzymes

20.4.2 Directionality and Complexity in the Origin and Early Evolution of Life

Knoll and Bambach [15] have identified six significant evolutionary steps or megatrajectories in the history of life, providing a broad pattern of the procession of life. Out of these six megatrajectories, the first two that are relevant to simple microscopic organisms represent (1) evolution from the origin of life to LUCA and (2) the metabolic diversification of bacteria and archaea (Fig. 20.2).

Life is the epitome of complexity at work (Fig. 4.1). Complexity is a measure of the degree of diversity within the system. It increases with an increasing number of constituent parts and linkages. We can numerically estimate degrees of complexity by counting the components of a system. For example, simple microbial life is composed of more than 25 million components compared to only ~1000 components in a volume in a personal computer [16]. The simplest known life is vastly more complex than any nonliving components that might have contributed to it (Table 20.1). Amazingly, such an intricate complex could arise from few, lifeless building blocks. The most obvious way to increase the complexity of prebiotic synthesis is by the acquisition of analog, hybrid, and digital information systems in tandem with biomolecules that separate living organisms from nonliving matter.

Table 20.1 A total number of molecules in a single *Escherichia coli* cell is found to have >25 million components packed in a volume of a cubic micrometer. In contrast, a personal computer has only ~ 1,000 components in a volume of perhaps 51. (Modified from Deamer [16])

Number of molecules in a single *Escherichia coli* cell	
Molecular components	Number of molecules
Kinds of proteins	1850 (mostly enyzmes)
Total number of proteins	2.36 million
RNA in ribosomes	18,700
Transfer RNA	205,000
Messenger RNA	Variable (depending on growth cycle)
DNA	One circular double helix
Lipid	22 million (mostly in cell membrane)
Lipopolysaccharide	1.2 million
Peptidoglycan	One (forms cell wall)
Glycogen	4360 (energy storage of cell)

20.5 Entropy and Life

The second law of thermodynamics posits that the entropy of the universe increases over time, and it also holds steadfast for large-scale systems. Entropy is a measure of the disorder of a system. It is a common experience that things or inanimate objects left to themselves eventually become disordered: lavas cool, mountains erode, glaciers melt, machines break, buildings crumble, gas diffuses through air, dead organisms rot, and

so on. In short, over time, energy tends to disperse or spread out. No matter how advanced our machines become, they can never wholly avoid wasting energy and running down. Thermodynamic entropy measures this loss of helpful energy, quantifying the dispersion of energy among the particles of a system and its diffusion throughout space. Its increase is a simple matter of probability. The system ultimately arrives at a maximum entropy state called 'thermodynamic equilibrium,' in which energy is uniformly distributed.

One essential difference between living and nonliving things is that the former tends to be much better at capturing energy from the environment and dissipating that energy as heat. Life is comprised of open systems nourished by free energy from the environment. Living things preserve their low levels of entropy over the course of their lifetime because they receive energy in the form of food and nutrients from their surroundings. Living organisms maintain a high degree of order at every level. Unlike closed systems, which settle into a state of thermal equilibrium, open systems maintain themselves far from equilibrium, maintaining low entropy in this 'steady state' characterized by continual flow and transformation of energy. In this way, living systems hold the second law's disordered outcome at bay while they live. In the thermodynamics of open systems, life exports the entropy that would otherwise lead to disorder and thus functional impairment [17, 18]. This trend of decreasing entropy in living systems may be called the biological arrow of time [14].

If we view the entire universe as a closed system, then it will appear that the entropy of the universe has reached its maximum. Life is a highly ordered system at every level, and when an organism develops and reproduces, the order of the system increases. As living organisms feed, develop, and grow, they appear to create this order out of raw materials that lack it. They must maintain their organization to function. Cells use stored energy to carry out nonspontaneous reactions that increase order or decrease entropy. However, in order to store energy in the first place, cells must perform spontaneous reactions, such as breaking down food molecules. These reactions increase the entropy of the universe. Overall, the increase in entropy by spontaneous reactions is greater than the decrease in entropy by the cell's nonspontaneous reactions. Even with the proliferation of life on Earth, the entropy in the universe continues to increase, and the second law is satisfied.

The second law can also be applied to prebiotic synthesis. In the energy-rich far-from-equilibrium environment of a hydrothermal vent, this trend is apparent in the biomolecules formed during biosynthesis. An increase in order and a reduction in entropy with the emergence of increasingly complex molecules with increased functionalities is compensated for when these biomolecules return part of this energy to the environment as entropy (heat and low-free-energy compounds, such as water and carbon dioxide (CO_2)). Driven by an external source of chemical energy and surrounded by a volcanic heat bath in the hydrothermal crater vent environment, these biomolecules gradually restructured themselves to dissipate the increasing energy. ATP served as the primary energy of the hydrothermal vents' currency and was used to build complex molecules. This could mean that biomolecules inexorably acquired, collected, and stored information—the key physical attribute associated with life—about their unpredictable environment. This information allowed those biomolecules to stay in a local environment far from equilibrium, under a constant flux of energy that drove organic reactions toward ever-increasing complexity. The self-assembly of complex biomolecules in a vent environment and their self-replication are also mechanisms by which a system dissipates increasing amounts of energy over time. These biomolecules also operated in a far-from-equilibrium state, whereas the second law succinctly states that isolated systems evolve toward thermodynamic equilibrium, the state of maximum entropy. Self-organizing processes develop in far-from-equilibrium states. The thermodynamics of nonequilibrium systems make the emergence of organized complex systems, such as life, much more likely far from equilibrium on prebiotic Earth than if the raw chemical ingredients were just lying about in hydrothermal vent environments. The forces underlying molecular self-assembly include covalent bonds, hydrogen bonds, electrostatic interactions, and the hydrophobic effect [3]. Self-assembling processes occurred spontaneously in vent environments where simple organic compounds in solution were organized into more complex systems.

Besides thermodynamic entropy, life also partakes in the informational entropy defined by information theory. Claude Shannon formalized the concept of informational entropy into three components: the message per se, the modalities of its transmission, and the semantic effects of its reception [19]. Entropy, according to Shannon's theory, is a measure of uncertainty about the identity of signals in a message. In our previous work [20], we suggested that life may be operationally defined as an information-processing system—a structural hierarchy of various functional units—acquired through evolution, i.e., the capacity for storing and processing the information necessary for its own accurate reproduction. The discovery of the genetic encoding carried by a DNA molecule and its mode of translation into protein structures secured the modern view of biology as information science.

Biological systems have embedded information structures that support their functions. Life avoids decay by importing, capturing, and storing information, or negative entropy, from its surroundings [21]. As Shannon's theory was adapted to bioinformatics, communication in this context was seen to require two components, a thermodynamic framework for the coding and transmission of signals and the semantic sense produced by the reception of the message.

The latter component bears no thermodynamic burden. In the transmission of genetic information, digital data from the base sequences of DNA are translated into functional proteins. For all life forms, the genotype's message encoded in DNA specifies the phenotype. Darwinian evolution is now primed to adapt the organism to its environment.

20.6 Life in Extreme Environments

The hyperthermophilic view of the origin of life led to the study of extremophiles and their environments. Extremophiles live at the edges of the biosphere, thriving in habitats that other terrestrial life-forms find intolerably hostile or even lethal. That microbial life is abundant and diverse in extreme environments has been one of the most important discoveries of recent times. The conditions prevailing in the more extreme habitable niches are taken to define a 'biological envelope' within which life can survive, grow, and evolve. Most extremophiles are bacteria and archaea, but a few are protists and microbial eukaryotes. They live in a diverse range of environments, exposed to extremes of heat, cold, pressure, darkness, toxins in volcanic vents, acid, alkaline, or salty niches, and deep rocks. For all their diverse properties, extremophiles have one thing in common: they need water to survive.

The biology of extremophiles suggests that the conditions under which life can arise and evolve elsewhere in the solar system and beyond are relatively extensive. NASA's Astrobiology Program has established expectations for the existence of at least microbial life in many other habitable places within and beyond our solar system. However, how do we detect it? How do we find those specific habitats on other planets where life may exist, and how do we get to them? So far, no matter how extreme the environment, all life on Earth has the same ancestry. Life could have happened differently on other planets or moons, making it challenging to figure out what to search for. The DNA-based life hosted on Earth may not be the same chain of life formed on other planets. Until we fully understand Earth's limits, it will be difficult to determine whether any planets or moons in the solar system could host any living beings. Extremophiles that are adapted to high levels of radiation may offer the most important clues to the possibility of extraterrestrial life in the inhospitable habitats of Mars, Jupiter's moon Europa, and Saturn's moons Titan and Enceladus. The discovery of extremophilic life lends significant support to the idea that microbial life might be widespread in the solar system and beyond. Simple microscopic organisms (megatrajectories 1 and 2, see Fig. 17.2) are the life forms most likely to be present in the universe [12]. Although there is still no direct evidence of life on other planets in our solar system or galaxy, recent NASA explorations have prompted optimism that habitable worlds exist throughout the Milky Way.

20.7 Life Here and Beyond

The study of life's origin has a wide range of implications for our attempts to answer the iconic question 'Are we alone?' Astrobiology addresses both the origin of life on Earth and our aspirations to become a space-faring civilization. Understanding the origin of life, the history of interactions between life and environment, and the wide range of extremophiles' habitats on Earth help us explore life in the solar system and exoplanets. Experimenting with life's beginnings on Earth, NASA has assembled a large body of knowledge about the chemical makeup of the cosmos. NASA is also analyzing the extraterrestrial dimensions of life by testing for habitability on missions to Venus, Mars, Jupiter's moon Europa, and Saturn's moons Titan and Enceladus. Any life elsewhere in the solar system would strengthen the expectation that life is commonplace in the universe.

20.8 Conclusions

Few human quests are as grand as the search for life's origin and possible distribution in the cosmos. How could something as exquisitely organized as a living cell come into being amidst a violent world ravaged by heavy meteoric bombardment from above and volcanism from below? The transition from nonlife to life was almost certainly a gradual process. To reconstruct the phases of this momentous event, we have investigated the plausible forms of primal elements present four billion years ago on primitive Earth. At that time, young Earth had cooled down sufficiently for water to form global oceans, where continental islands, dotted with hydrothermal crater lakes, became the cradles for prebiotic synthesis, the simmering cauldrons where organic compounds delivered by meteoritic infall, and geochemical synthesis accumulated, and the components of life began to brew. Some organic compounds were monomers that polymerized, whereas others were amphiphilic molecules that spontaneously assembled into membranous structures encapsulating the monomers and polymers to form myriads of protocells. Living cells arose and preserved their internal order by taking free energy from their surroundings in the form of chemical nutrients and, later, sunlight and then returning an equal amount of energy as heat and entropy to their surroundings. As Schrödinger once put it, life involves an increase in negative entropy. Therefore, life's origin and development may be understood, wherever it may arise, as a local transition from chaos to order. Life exists on Earth

because the universal law of entropy drives matter into forms of complex organization that eventually fall into self-maintaining and self-reproducing cellular bodies.

Life comprehends the interaction of all the inanimate components of this whole system. In other words, nothing is alive in a cell, except the whole system itself. Crucial cell components, such as the plasma membrane, nucleic acids, and proteins, though sophisticated and organized, are not alive, and they obey the same physical and chemical laws as inanimate systems. Life is an emergent property whose systemic behavior cannot be understood or predicted from its individual parts or cell components alone. Like life, order is an emergent property out of chaos. Emergent properties necessarily arise from the collective interactions of the parts within a more extensive system. For instance, living systems are composed of biopolymers, such as nucleic acids and proteins, which interact within a lipid bilayer membrane with three primary functions: containment, transport of nutrients, and energy transduction. Nucleic acids such as mRNA and DNA have a unique ability to store and transmit digital information. Similarly, protein enzymes have a unique ability to catalyze metabolic reaction rates. Genetic and catalytic polymers formed a symbiotic relationship that increased the efficiency of both biopolymers. Both genes and proteins are manufactured by molecular machines. Life is made of manufactured objects, but nonlife is made of spontaneous objects. Life began when a few of the immense numbers of protocells incorporated a symbiotic cycle involving genetic information and catalytic functions.

Viruses played critical roles in the origin of life. They became active participants in the protein/RNA world by recurrent infection of protocells by horizontal gene transfer (HGT) and horizontal enzyme transfer (HET) and by exploiting their ribosomes for reproduction. However, donating the RdRp enzyme to protein/RNA protocells gave rise to primordial protocellular life. The next stage in viral evolution was the emergence of a primitive retrovirus with a new kind of reverse transcriptase (RT) enzyme that could turn its protein/RNA protocell into a protein/DNA protocell. With persistent infection, DNA viruses slowly transfer to protocells, not only their core replication enzymes by HET but also their DNAs by HGT. Thus, began the 'DNA world' when DNA replaced RNA as the major genome of protocells. With the invention of transcription and translation by DNA, the first cellular life emerged.

The emergence of DNA and first cells gave rise to a new mechanism for generating genetic diversity that accelerated evolution. Yet, the rules guiding the process of hierarchical self-assembly remained virtually unchanged.

References

1. Forterre P (2006) The origin of viruses and their possible roles in major evolutionary transitions. Virus Res 117:5–16
2. Bansho Y, Furubayashi T, Ischihasi N, Yomo T (2016) Host-parasite oscillation dynamics and evolution of compartmentalized RNA replication system. Proc Natl Acad Sci U S A 113:4045–4050
3. Monod J (1970) Chance and necessity: an essay on the natural philosophy of modern biology. Knopf, New York
4. Davies P (2016) Many planets, not much life. Sci Am 315:8. https://doi.org/10.1038/scientificamerican09167-8
5. Crick F (1981) Life itself: its origin and nature. Simon and Schuster, New York
6. Ward PD, Brownlee D (2001) Rare Earth: why complex life is uncommon in the universe. Copernicus Press, Newton
7. De Duve C (2011) Life as a cosmic imperative? Phil Trans R Soc A 369:620–623
8. Gould SJ (2002) The structure of evolutionary theory. Belknap Press, Harvard University, Cambridge, MA
9. Morowitz HJ (1992) Beginnings of cellular life: metabolism recapitulates biogenesis. Yale University Press, New Haven
10. Bird RJ (2003) Chaos and life. Columbia University Press, New York
11. Laursen L (2013) Predictable evolution trumps randomness of mutations. Nature. https://doi.org/10.1038/nature.2013.12459
12. Barbieri M (2018) What is information? Philos Trans R Soc A 74. https://doi.org/10.1098/rsta.2015.0060
13. Chatterjee S, Yadav S (2022) The coevolution of biomolecules and prebiotic information systems in the origin of life: a visualization model for assembling the first gene. Life 12:834. https://doi.org/10.3390/life12060834
14. Walker I (1972) Biological memory. Acta Biotheor 21:203–235
15. Knoll AH, Bambach RK (2000) Directionality in the history of life: diffusion from the left wall or repeated scaling of the right. Paleobiol 26(4):1–14
16. Deamer D (2009) On the origin of systems. EMBRO Rep 10:51–54
17. Schrödinger E (1944) What is life? The physical aspects of living cell. Cambridge University Press, Cambridge
18. Prigogine I (1980) From being to becoming. W. H. Freeman, New York
19. Shannon CE (1948) A mathematical theory of communication. Bell Sys Tech J 27:379–423 & 623–656
20. Chatterjee S, Yadav S (2019) The origin of prebiotic information system in the peptide/RNA world: a simulation model of the evolution of translation and the genetic code. Life 9:25. https://doi.org/10.3390/life9010025
21. Davies P (1999) The fifth miracle. Simon & Schuster, New York

Index

S. Chatterjee, *From Stardust to First Cells*, https://doi.org/10.1007/978-3-031-23397-5