

Akihiko Yamagishi · Takeshi Kakegawa
Tomohiro Usui *Editors*

Astrobiology

From the Origins of Life to the Search for
Extraterrestrial Intelligence

 Springer

Astrobiology

Akihiko Yamagishi
Takeshi Kakegawa • Tomohiro Usui
Editors

Astrobiology

From the Origins of Life to the Search
for Extraterrestrial Intelligence

 Springer

Editors

Akihiko Yamagishi
Department of Applied Life Sciences
Tokyo University of Pharmacy
and Life Sciences
Tokyo, Japan

Takeshi Kakegawa
Graduate School of Science
Tohoku University Geosciences
Miyagi, Japan

Tomohiro Usui
Department of Solar System Sciences
Institute for Space and Astronautical
Science, Japan Aerospace Exploration
Agency, Sagamihara, Kanagawa, Japan

ISBN 978-981-13-3638-6 ISBN 978-981-13-3639-3 (eBook)
<https://doi.org/10.1007/978-981-13-3639-3>

Library of Congress Control Number: 2019930834

© Springer Nature Singapore Pte Ltd. 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Astrobiology is the multidisciplinary field spreading from astronomy, planetary science, geology, biology, and sociology. Recent progress in these scientific fields, including the findings of organic compounds on Mars and icy satellites and the detection of more than 3500 exoplanets, has promoted us to consider the possible presence of life in space. To explore the solar system, many exploration programs are in schedule and more in consideration. Signature of life including the presence of organic compounds and spectroscopic characteristics are targeted to be searched in these celestial bodies.

Our exploration area is expanded toward outside the solar system. Telescopes with high resolution, sensitivity, and contrast are scheduled or considered to be constructed on the ground or in space. The exoplanets are targeted to search for the signature of life such as of the oxygen, ozone, and photosynthetic pigments.

To conduct the efficient search for life in the universe, it is essential to have comprehensive knowledge on the life on Earth. Current studies on astronomy and planetary science revealed that the Earth was positioned in a miracle place to be habitable when the solar system started. The early Earth had suitable hydrosphere and atmosphere to promote chemical evolution for origin of life. Afterward, the environments of the Earth changed dynamically by meteorite impacts, plate tectonics, global volcanisms, change of brightness of sun, etc. Knowledge on the origin and evolution of life with the close interaction with the change in Earth's environments is crucial to understand why we are here on the Earth.

To search for life, we also need to know what is life. We have to design the search program with better knowledge on what is going to be searched, where to search, and how to search. The key elements are organic compounds which enable the emergence of life and thermodynamic nonequilibrium which sustains life. The consequence of the emergence of life especially of photosynthesis that changed the atmosphere can be the target of life search, too.

To search for another target, intelligent life, we also need to know how intelligence emerges and to understand the evolution of human being. Search for extraterrestrial intelligence (SETI) started more than a half century ago. From the first attempt to search for intelligent life, the total capability of radio telescope has

elevated 10^{26} -folds. The Square Kilometre Array (SKA) is the radio telescope system consisting of thousands of radio telescopes with the total receiver size of approximately one square kilometer. Construction of SKA has started at two sites, one in Australia and the other in South Africa. SKA is going to be used to approach significant scientific targets including the search for intelligent life. SKA has potential to find signs of intelligent life, though the area is still limited to the exoplanets in the vicinity of our solar system.

This book is dedicated to show the comprehensive knowledge on these fields with the sufficient coverage of most recent finding as well as the introduction to the basic understanding of the vast field of astrobiology. This book will be useful not only for students but also for the scientists who want to know the progresses in the fields different from their own disciplines. This book will also give the reader with sufficient introduction and show the way to the deeper knowledge of each field.

Tokyo, Japan
Miyagi, Japan
Kanagawa, Japan

Akihiko Yamagishi
Takeshi Kakegawa
Tomohiro Usui

Contents

Part I Introduction to Astrobiology

- 1 What Is Astrobiology?** 3
Akihiko Yamagishi

Part II Physics and Chemistry from Space to Life

- 2 Prebiotic Complex Organic Molecules in Space** 11
Masatoshi Ohishi
- 3 Chemical Interactions Among Organics, Water, and Minerals
in the Early Solar System** 23
Hikaru Yabuta
- 4 Prebiotic Synthesis of Bioorganic Compounds by Simulation
Experiments** 43
Kensei Kobayashi
- 5 RNA Synthesis Before the Origin of Life** 63
Yoshihiro Furukawa

Part III History of Life Revealed from Biology

- 6 RNA World** 77
Shotaro Ayukawa, Toshihiko Enomoto, and Daisuke Kiga
- 7 The Common Ancestor of All Modern Life** 91
Satoshi Akanuma
- 8 Eukaryotes Appearing** 105
Shin-ichi Yokobori and Ryutaro Furukawa
- 9 Color of Photosynthetic Systems: Importance of Atmospheric
Spectral Segregation Between Direct and Diffuse Radiation** 123
Atsushi Kume

10	Evolution of Photosynthetic System	137
	Satoshi Hanada	
11	Cosmolinguistics: Necessary Components for the Emergence of a Language-Like Communication System in a Habitable Planet	153
	Kazuo Okanoya	
12	Evolution of Intelligence on the Earth	167
	Mariko Hiraiwa-Hasegawa	
Part IV History of the Earth Reveiled from Geology		
13	Formation of Planetary Systems	179
	Shigeru Ida	
14	Evolution of Early Atmosphere	197
	Hidenori Genda	
15	Biogenic and Abiogenic Graphite in Minerals and Rocks of the Early Earth	209
	Takeshi Kakegawa	
16	Cellular Microfossils and Possible Microfossils in the Paleo- and Mesoarchean	229
	Kenichiro Sugitani	
17	Great Oxidation Event and Snowball Earth	261
	Eiichi Tajika and Mariko Harada	
18	End-Paleozoic Mass Extinction: Hierarchy of Causes and a New Cosmoclimatological Perspective for the Largest Crisis	273
	Yukio Isozaki	
19	Mass Extinction at the Cretaceous–Paleogene (K–Pg) Boundary	303
	Teruyuki Maruoka	
Part V Search for Life in Solar System and Extra Solar System		
20	Limits of Terrestrial Life and Biosphere	323
	Ken Takai	
21	What Geology and Mineralogy Tell Us About Water on Mars	345
	Tomohiro Usui	
22	Atmosphere of Mars	353
	Hironmu Nakagawa	
23	The Search for Life on Mars	367
	Yoshitaka Yoshimura	

24 Active Surface and Interior of Europa as a Potential Deep Habitat 383
Jun Kimura

25 Enceladus: Evidence and Unsolved Questions for an Ice-Covered Habitable World 399
Yasuhito Sekine, Takazo Shibuya, and Shunichi Kamata

26 Astrobiology on Titan: Geophysics to Organic Chemistry 409
Hiroshi Imanaka

27 Panspermia Hypothesis: History of a Hypothesis and a Review of the Past, Present, and Future Planned Missions to Test This Hypothesis..... 419
Yuko Kawaguchi

28 Extrasolar Planetary Systems..... 429
Motohide Tamura

29 How to Search for Possible Bio-signatures on Earth-Like Planets: Beyond a Pale Blue Dot 441
Yasushi Suto

30 SETI (Search for Extraterrestrial Intelligence) 451
Hisashi Hirabayashi

31 Possible Cultural Impact of Extraterrestrial Life, if It Were to Be Found..... 461
Junichi Watanabe

Part I
Introduction to Astrobiology

Chapter 1

What Is Astrobiology?



Akihiko Yamagishi

Abstract Astrobiology is a multidisciplinary scientific field encompassing biology, chemistry, physics, geology, planetary science and astronomy. Scientists participating in this field are interested in trying to answer fundamental questions of life. These questions may include the search for extraterrestrial life through the discovery of primitive life forms such as bacteria or by detecting other intelligent beings in the universe. The basic premise of this research is to deepen knowledge of ourselves as human beings.

Astrobiologists are required to apply a scientific approach to assessing the potential for extraterrestrial life forms and developing methods for their discovery. There are possible locations for extraterrestrial life to thrive on Mars and icy satellites. Another target in the search for life is extrasolar planets. However, we require a method to search for life outside of Earth, but based on our current knowledge of life on Earth.

Keywords Origin · Evolution · Distribution · Exploration · Life

1.1 Introduction

Astrobiology is a multidisciplinary scientific field encompassing biology, chemistry, physics, geology, planetary science and astronomy. Scientists participating in this field are interested in trying to answer fundamental questions of life such as “Where did we come from? Are we alone? Where are we going?” (Bertka 2009), or “What is life? What is the course of life? Who are we?” (Sullivan and Barross 2007). These questions may be more specific: “How did life originate and diversify? How does life co-evolve with a planet? Does life exist beyond the Earth? What is the future of life on the Earth?” (Cockell 2015). These are the questions that motivate

A. Yamagishi (✉)
Department of Applied Life Sciences, Tokyo University of Pharmacy and Life Sciences,
Tokyo, Japan
e-mail: yamagish@toyaku.ac.jp

the search for extraterrestrial life through detection of very primitive forms of life, such as bacteria, or the presence of other intelligent beings.

These questions are motivated by a fundamental need to know ourselves as human beings and terrestrial life but are challenging to answer because there are no other life forms or intelligent beings known to us for comparison. To answer these fundamental questions, we need to find another form of life. However, our search for other life forms needs to be guided by the answers to these fundamental questions. This poses one of the challenges that must be addressed in astrobiology.

There are other textbooks focused on astrobiology, including an introductory guidebook (Catling 2013) and a textbook for an undergraduate course with an accompanying web site (Cockell 2015). There are good overviews covering planets, life and intelligence (Ulmschneider 2006), a multi-authored review book encompassing the search for extraterrestrial intelligence and alien biochemistries (Sullivan and Baross 2007) and one with even more emphasis on philosophical and ethical perspectives (Bertka 2009; Vakoch 2013). As opposed to these textbooks, this book was intended to be a concise but comprehensive handbook, not a textbook describing basic concepts in detail but sufficient for a scientist outside of his profession to comprehend a current overview of the field.

In this book, each chapter will summarize the current state of the continuously developing field of astrobiology. In this broad field, students and senior scientists must learn a different way of thinking as well as be able to understand different terminology from the divergent sciences that make up astrobiology. Each chapter will summarize the relevant discoveries in each field to provide readers with an overview of current research with sufficient references for more in-depth understanding.

1.2 Why Astrobiology Now?

The term astrobiology was introduced in 1995 by Wes Huntress, based at NASA's headquarters in Washington, D.C. (Catling 2013). NASA was trying to develop Mars missions to locate water and then search for signatures of life. Since then, explorations of the solar planets have revealed the novel appearance of solar system bodies. Mars was considered a dead planet that lost geological activity and most of its atmosphere, but the explorations of Mars revealed aspects of the planet indicating it is still at least partially active (Chaps. 21, 22 and 23). Other active environments such as hydrothermal fields have been found under the frozen oceans of icy satellites of Jupiter and Saturn (Chaps. 24 and 25). These are all possible locations for extraterrestrial life to thrive.

Another target in the search for life is extrasolar planets. As reviewed in Chap. 28, more than 3500 extrasolar planets have been found, and the number is increasing every year. The extrasolar planets include those of a similar size as Earth and those within the habitable zone, which is the estimated distance from the central star allowing for an expected temperature permitting the presence of liquid water and oceans. A possible habitable planet was found in the closest star to our solar system, which is only at about 4 light-year distance from Earth.

These extrasolar planets are current targets in the search for life and prompt us to consider what life is, so that we can develop methods to search for it. We experience a dilemma when we consider the form that extraterrestrial life is expected to take. If we tend towards fact-based predictions, extraterrestrial life should resemble terrestrial life. On the contrary, imagination lets us imagine extraterrestrial life forms highly divergent from terrestrial life forms. These contradictions facilitate development of new avenues of life science, where scientists have started the scientific reevaluation of knowledge on life, planets and the universe.

1.3 Why Astrobiology Is Needed

There are several aspects of astrobiology that must be addressed. One is the science versus fantasy of extraterrestrial life. We know of many fiction movies involving extraterrestrial intelligent species attacking and invading the Earth, either via infectious disease or by supernatural beings. Of course, scientists are aware of the differences between science and fantasy but also realize that knowledge on extraterrestrial life is lacking. Astrobiologists must reinforce a scientific approach to predicting possible extraterrestrial life forms and styles.

There is also a need for further development of scientific approaches to the search for life. We must develop techniques for searching for life beyond Earth that are based on our current knowledge of life on Earth.

1.4 Textbook Overview

Astrobiology is a scientific field that studies life from a universal point of view, not only related to life alone nor limited to terrestrial life alone. The vast field of astrobiology is divided into four sections in this book. In Part II, we will follow the synthesis and accumulation of organic compounds before the origin of life on the Earth. Terrestrial life consists of 70% water with the remainder mostly organic compounds (Table 1.1). Organic compounds were found in a molecular cloud in the Galaxy, and they must have been transported to primitive Earth before the origin of

Table 1.1 Molecular composition of *Escherichia coli* cell (Watson 1976)

	Composition (%)
Water	70
Protein	15
Nucleic acid DNA	1
RNA	6
Lipid	3
Carbohydrate	4
Mineral	1

life. Understanding the processes of synthesis, modification and transport of organic compounds is the target of research in astrobiology. RNA, one of the most important organic compounds, may have been produced on Earth.

In Part III of this book, we will follow the origin and evolution of terrestrial life and intelligence. We will summarize major evolutionary time points including RNA as the origin of life, the common ancestor, origin of eukaryotes, origins of photosynthesis, language and intelligence. We will follow the co-evolution of life and the planet Earth from a biological perspective.

The formation and evolution of planets are also the target of research in astrobiology and will be reviewed in Part IV. We know the Earth is a rocky planet with approximately 70% of its surface covered by ocean. We know that the presence of an ocean seems to be necessary for life to emerge. However, we do not know how and from where adequate amounts of water came to primitive Earth. The current knowledge on how planet systems and early atmosphere emerged will be reviewed. Geological records, including biological fossils, are important information that allows us to reconstruct the history of Earth. The fossil record of carbon and cellular fossils will also be explained. Major impacts on the biological history on Earth, including the great oxidation event and mass extinction events, will be addressed.

In the Part V, the current status and future direction of exploration of life will be addressed. The limits of life on Earth will provide a tentative guideline where to search for other forms of life. The main targets in the solar system are Mars and icy planets around Jupiter and Saturn. The possible transfer of life between planets may alter scientists' approach to searching for life; this hypothesis is called Panspermia. The search for extrasolar planetary system, life and intelligence will be addressed in the final chapters, including a discussion on the possible implications once they are found.

1.5 Conclusion

Astrobiology is a multidisciplinary field of research addressing fundamental questions of life such as how and where life emerged. How has the Earth emerged and co-evolved with life? Less often asked, but of equal importance, is "Why has life emerged and evolved in the universe?". These questions all try to understand the emergence of life and consequently further understanding of ourselves. By knowing our universe, planet and ourselves, it will be possible to predict and prepare for the future of our world.

References

- Bertka CM (ed) (2009) Exploring the origin, extent, and future of life: philosophical, ethical, and theological perspectives. Cambridge University Press, Cambridge
- Cataling DC (2013) Astrobiology: a very short introduction. Oxford University Press, Oxford
- Cockell CS (2015) Astrobiology: understanding life in the universe. Wiley, Chichester
- NASA Home page: <https://astrobiology.nasa.gov/about/>
- Sullivan WT III, Baross JA (eds) (2007) Planets and life: the emerging science of astrobiology. Cambridge University Press, Cambridge
- Ulmschneider P (2006) Intelligent life in the Universe. Springer, Berlin
- Vakoch V (ed) (2013) Astrobiology, history, and society: life beyond Earth and the impact of discovery. Springer, Berlin
- Watson JD (1976) Molecular biology of the gene, 3rd edn. Benjamin, Menlo Park, p 69

Part II
Physics and Chemistry from Space to Life

Chapter 2

Prebiotic Complex Organic Molecules in Space



Masatoshi Ohishi

Abstract As of 2017, about 200 complex organic molecules have been detected in interstellar molecular clouds. It was 1969 when the first organic molecule in space, H_2CO , was discovered. Since then many organic molecules were discovered by using the NRAO 11 m (upgraded later to 12 m), Nobeyama 45 m, IRAM 30 m, and other highly sensitive radio telescopes as a result of close collaboration between radio astronomers and microwave spectroscopists. It is noteworthy that many well-known organic molecules such as CH_3OH , $\text{C}_2\text{H}_5\text{OH}$, $(\text{CH}_3)_2\text{O}$, and CH_3NH_2 were detected in the 1970s. It is thought that organic molecules are formed on surfaces of cold dust particles in a molecular cloud and then are evaporated by the UV photons emitted from a star inside the molecular cloud.

Organic molecules are known to exist in star-forming regions and in protoplanetary disks where planets are formed. Therefore it was a natural consequence that astronomers considered a relationship between organic molecules in space and the origin of life. Several astronomers challenged to detect glycine and other prebiotic molecules without success. ALMA is expected to detect such important materials to further examine the “exogenous delivery” hypothesis of organic molecules.

In this chapter I summarize the history of the searches for complex organic molecules in space together with difficulties in observing very weak signals from larger molecular species.

Keywords Interstellar molecules · Radio astronomical observations · Dust particles · Exogenous delivery

M. Ohishi (✉)
National Astronomical Observatory of Japan, Tokyo, Japan
e-mail: masatoshi.ohishi@nao.ac.jp

2.1 Introduction

It is well-known that our solar system was formed from its natal molecular cloud 4.6 billion years ago. After the formation of Earth and other planets in our solar system, these planets underwent frequent impacts by smaller bodies for several hundred million years (so-called late heavy bombardment). Though life on Earth may have emerged during or shortly after or before this heavy bombardment phase, as early as about 3.9 billion years ago, discussions are continuing on its exact timing. In 2017 a paper was published that claims a discovery of early trace of life in sedimentary rocks at 3.95 billion years ago (Tashiro et al. 2017; see also Chap. 15). It means that life on Earth existed about 4 billion years ago, i.e., only 600 million years after the formation of Earth.

It is generally accepted that origin of life was the result of chemical evolution of organic molecules on the primordial Earth. Thus, an important issue in astrobiology is where prebiotic organic molecules are formed, terrestrial or extraterrestrial. After the famous Urey-Miller's experiment, many researchers believed that the primary formation site of prebiotic organic molecules was the surface of Earth under reducing atmosphere (see Chap. 4). Recent modelling of the Earth's early atmosphere suggests more neutral conditions which preclude the formation of significant amount of prebiotic organic molecules. The situation, in turn, lead people to consider another possibility: delivery of extraterrestrial prebiotic organic molecules through comets, asteroids, meteorites, and interplanetary dust particles (the exogenous delivery hypothesis). One research suggested that extraterrestrial organic compounds may be more abundant by three orders of magnitude than their terrestrial formation (Ehrenfreund et al. 2002). It is likely that a combination of these sources contributed to the building blocks of life on the early Earth (see Fig. 2.1). What is certain is that once life emerged on the primordial Earth, it was capable to adapt quickly to the surrounding environment for its survival through finding shelter from the UV photons and energy source. This continuing process led to complex metabolic life and even our own existence.

In this chapter, I will review prebiotic organic molecules in space, which may be brought to the early Earth. If the exogenous delivery hypothesis works, similar process would occur even in other extrasolar Earth-like planets.

2.2 Molecules in Space

2.2.1 *Molecules Observed in Space*

Molecules exist in a variety of physical conditions in the Universe. It was in 1940 when unidentified ultraviolet (UV) absorption lines were attributed to CH and CN in diffuse interstellar clouds (McKellar 1940). This was the first report of molecules in space.

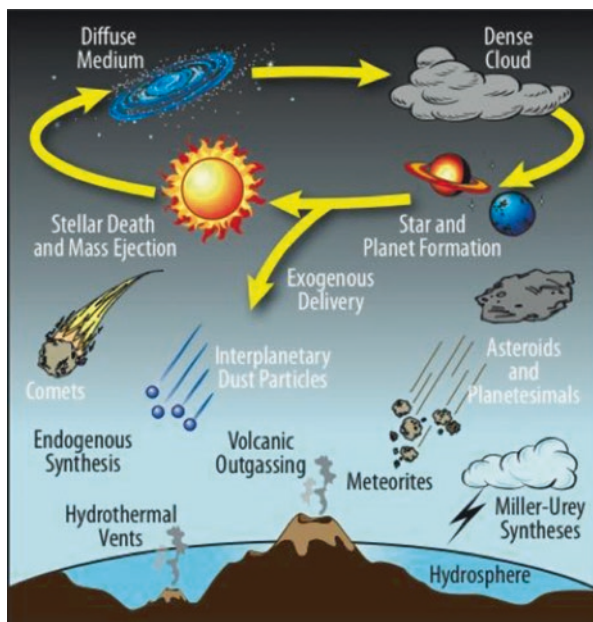


Fig. 2.1 A schematic diagram on the origin of prebiotic organic molecules. (Courtesy by Dr. Amie Elsila (Deamer et al. 2002))

In the early 1930s, radio astronomy opened new eyes to the universe. In 1963 the OH molecule was discovered by a radio telescope (Weinreb et al. 1963). As of December 2017, some 200 molecules were detected or reported, primarily by means of radio astronomical observations, in interstellar clouds, circumstellar envelopes, and even external galaxies. The smallest and the lightest molecule is H_2 . A list of molecules in space can be found in, e.g., the Cologne Database for Molecular Spectroscopy (2001).

1969 was the year when the first organic polyatomic molecule, H_2CO , was detected toward many galactic and extragalactic sources. Prior to the detection of H_2CO , the first “large” millimeter-wave telescope, the 11 m radio telescope of the National Radio Astronomy Observatory (NRAO) in the USA, started its operation in 1968. The NRAO 11 m radio telescope was one of pioneering telescopes (later, this telescope was upgraded to 12 m); it discovered CO, CH_3CN , HCN, HNC, $(CH_3)_2O$, C_2H , C_2H_5OH , NH_2CN , C_2H_5CN , H_2CCO , C_3N , and HNCN in the 1970s. It was in the 1970s when other countries joined the “molecular hunting.” Australia used the Parkes 64 m radio telescope and detected H_2CS , CH_2NH , CH_2CHCN , and $HCOOCH_3$. Japan constructed the 6 m millimeter-wave telescope and succeeded in discovering CH_3NH_2 . Detailed detection history can be found in the reference (Ohishi 2016). Spectral line surveys are powerful means in detecting new molecules in space. Figure 2.2 shows an example of a spectral line survey observation toward a cold, dark molecular cloud, Taurus Molecular Cloud 1 (TMC-1) conducted by

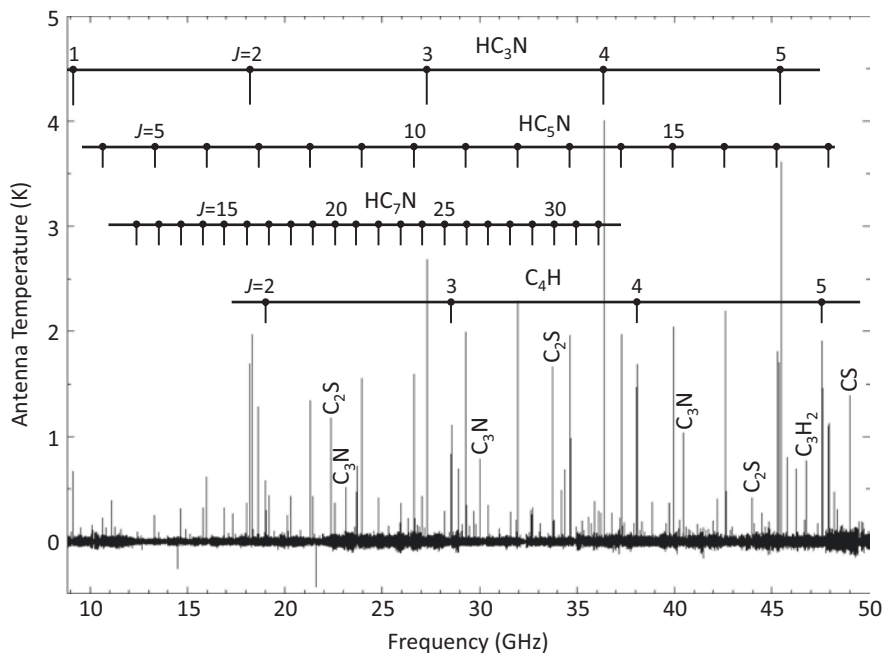


Fig. 2.2 Observed spectrum from 8.8 to 50.0 GHz toward TMC-1 by the Nobeyama 45 m radio telescope. The ordinate, the antenna temperature, is usually used in radio astronomy for expressing energy of received radio waves; it is known that the Boltzmann constant (k) times temperature (T) is equal to energy. The frequency resolution was 37 kHz, and the total number of data bins is 2 million. (This figure is modified from Fig. 1 in Ohishi (2016)). A part of identified molecular species are shown with their chemical formulas. The horizontal bars with the quantum number J , which denotes the total angular momentum of the molecular rotations, show spectral patterns by linear molecules (equal frequency intervals) such as HC_3N , HC_5N , and HC_7N

using the 45 m radio telescope at the Nobeyama Radio Observatory, the National Astronomical Observatory of Japan (Kaifu et al. 2004). We are able to see clear spectral patterns by linear molecules (with equal frequency intervals) such as HC_3N , HC_5N , and HC_7N , in Fig. 2.2. The NAOJ team succeeded in detecting 17 new molecules in space through the spectral line survey observations.

It should be stressed that many organic molecules were already detected in the 1970s. It is well-known that CH_3OH is widespread in a variety of sources; CH_3OH masers are commonly observed toward young star-forming regions. Large organic molecules, such as CH_3CN , CH_3CHO , $(\text{CH}_3)_2\text{O}$, and NH_2CHO , are usually observed toward dense and hot regions with the number density of molecular hydrogen ($n(\text{H}_2) > 10^{7-8} \text{ cm}^{-3}$ and the gas kinetic temperature $> 200 \text{ K}$).

2.2.2 *Classification of Molecules in Space*

Molecules in space can be classified into several categories: simple molecules, molecular ions, radicals, ring molecules, and stable molecules. Simple molecules can be seen in a variety of sources. Among them, the most abundant molecule is H_2 . Indeed, around 99.99% of molecules in space are H_2 . The next abundant molecules are H_2O and CO whose relative abundances to H_2 are about 10^{-4} . H_2O , CO , and CO_2 are ubiquitous molecules in the solid phase (Herbst and van Dishoeck 2009); these molecules in dust mantles are often observed in absorption toward bright infrared (IR) sources. Molecular ions and radicals are characteristic for molecules in space (Herbst and van Dishoeck 2009). Under the terrestrial physical conditions where the mean free time of these molecules is the order of milliseconds, such molecular ions or radicals react with other atom or molecule and are converted into other species immediately after they are formed. Thus, they are often called as “short-lifetime” molecules. However, under the interstellar and circumstellar conditions where the mean free time is typically of the order of a year, such “short-lifetime” molecules can survive for a long time. It is now known that molecular ions play an important role in the gas phase where “ion-molecule” reactions can successfully explain the formation of many major molecules in space (Herbst and van Dishoeck 2009).

Large “complex organic molecules” with six or more atoms (hereafter COMs; Herbst and van Dishoeck 2009), including prebiotic molecules, are the primary topic of this chapter; they belong to “stable molecules.” Many of large COMs are closed-shell molecules, such as CH_3OH and $\text{C}_2\text{H}_5\text{OH}$, which exist stably under the standard terrestrial conditions. Past studies showed that many of large organic species are formed on cold (10–15 K) dust grains through the hydrogenation (addition of hydrogen atoms) to small molecules which are adsorbed from the gas phase to the surface of the dust grains (Herbst and van Dishoeck 2009). For example, gas-phase CO is adsorbed onto dust grains. Since the hydrogen atoms on the dust surface can move quickly through the “tunneling effect,” CO is converted to H_2CO and ultimately to CH_3OH (Herbst and van Dishoeck 2009). When a star is formed in the center of a molecular cloud core, the solid-phase CH_3OH molecules may be heated by the UV photons from the central star and evaporated back to the gas phase; such gas-phase large COMs can be observed by telescopes. Since prebiotic molecules are categorized into COMs, a large number of prebiotic molecules are thought to be formed on the surfaces of dust particles (Herbst and van Dishoeck 2009).

2.3 Prebiotic Organic Molecules in Space

2.3.1 *Why Prebiotic Organic Molecules Are Made of H, C, N, and O?*

In Sect. 2.2.1, I described that many organic molecules in space have been known since 1970s, and there are a variety of molecules which can be related to “life.” We know that the building blocks of life contain primarily H, C, N, and O. There would be a natural reasoning why molecules containing these four elements are abundant in space.

The cosmic elemental abundances are well-known in astronomy (see Table 2.1). It is clear that H, C, N, and O are the four most abundant elements in the Universe, except for He. This is the primary reason why water (H₂O) in space is the second most abundant molecule next to H₂ (see Sect. 2.2.2). With a similar consideration, it would be easily understood why carbon-bearing species, i.e., organic molecules, can be abundant in space. Helium is not contained in building blocks of life, which is chemically inactive.

2.3.2 *Prebiotic Organic Molecules of the Greatest Importance in Space*

In this subchapter, I will show a few prebiotic molecules discovered in space, which would have the greatest importance to astrobiology. Other prebiotic molecules are also important.

Table 2.1 Elemental abundance of atoms (relative to Si) from Hydrogen to Iron

Element	Abundance	Element	Abundance
Hydrogen (H)	28,000	Silicon (Si)	1
Helium (He)	2700	Phosphine (P)	0.008
Lithium (Li)	0.0000004	Sulfur (S)	0.45
Beryllium (Be)	0.0000004	Chloride (Cl)	0.009
Boron (B)	0.000011	Argon (Ar)	0.1
Carbon (C)	10	Potassium (K)	0.0037
Nitrogen (N)	3.1	Calcium (Ca)	0.064
Oxygen (O)	24	Scandium (Sc)	0.000035
Fluorine (F)	0.001	Titanium (Ti)	0.0027
Neon (Ne)	3	Vanadium (V)	0.00028
Sodium (Na)	0.06	Chromium (Cr)	0.013
Magnesium (Mg)	1	Manganese (Mn)	0.0069
Aluminum (Al)	0.083	Iron (Fe)	0.9

Taken from Chronological Scientific Tables (2017)

Glycolaldehyde (CH₂OHCHO)

An epoch-making event in astrobiology was the discovery of glycolaldehyde in 2000 toward Sagittarius B2 (Sgr B2), a giant molecular cloud in the central region of the Milky Way galaxy (Hollis et al. 2000). The paper was titled “Interstellar Glycolaldehyde: The First Sugar.” In general sugar is an important constituent of life: ribose (C₅H₁₀O₅) is a pentose class sugar monomer, and its modified form, deoxyribose, is known to be the backbone of the DNA. The first detection report was based on observed data by using the NRAO 12 m, with relatively poor signal-to-noise ratios. In 2004 a confirmation paper was published by using a new 100 m radio telescope, the Green Bank Telescope (GBT) (Hollis et al. 2004). Glycolaldehyde was detected around a solar-type young star, IRAS16293-2422, through Atacama Large Millimeter Array (ALMA) observations (Jørgensen et al. 2012). It was suggested that UV photochemistry of a CH₃OH-CO mixed ice on dust surfaces, which has undergone mild heating, would form glycolaldehyde. Followed by these detections, glycolaldehyde has been reported detected in other massive star-forming regions and solar-type star-forming regions.

According to the detection report (Hollis et al. 2000), the generic form of sugar can be expressed as (H₂CO)_n ($n > 2$). Because glycolaldehyde has the form of (H₂CO)₂, it was regarded as the simplest sugar in space. This discovery had led astronomers to seriously search for other “prebiotic molecules” in space. It was unfortunate that the correct generic form of sugar is (H₂CO)_n ($n > 3$), not 2!, i.e., glycolaldehyde is not a sugar.

n- and *i*-Propyl Cyanide (*n*- and *i*-C₃H₇CN)

It is known that meteorites found on Earth contain a variety of COMs including more than 80 amino acids (Elsila et al. 2007). Since meteorites are formed in protoplanetary nebula which are formed in interstellar medium, the amino acids and their precursors would have an interstellar origin. In this regard, it is crucial for astrobiology to understand how complex organic species can be formed in the interstellar medium.

In 2009 the first detection of normal-propyl cyanide (*n*-C₃H₇CN), the largest organic molecule in this source, was reported toward Sgr B2 by using the 30 m radio telescope operated by the Institut de Radioastronomie Millimétrique (IRAM) (Belloche et al. 2009). Propyl cyanide is the smallest alkyl cyanide, which has a few isomers: a straight-chain isomer, normal-propyl cyanide (also known as butyronitrile or 1-cyanopropane), and a branched-chain isomer, iso-propyl cyanide (*i*-C₃H₇CN, also known as iso-butyronitrile or 2-cyanopropane). Five years after the detection of *n*-C₃H₇CN, *i*-C₃H₇CN was detected for the first time toward SgrB2 by using ALMA (Belloche et al. 2014). Amazingly the abundance of *i*-C₃H₇CN was 40% of *n*-C₃H₇CN, suggesting that branched carbon-chain molecules may be generally abundant in the interstellar medium. The detection of *i*-C₃H₇CN further suggests the presence of amino acids in the interstellar medium, for which such branched-chain structure is a key characteristic.

Propylene Oxide (CH₃CHCH₂O)

The origin of homochirality in biological molecules, especially the use of only the L-amino acids and D-sugars, is a long-standing question to be solved in astrobiology. A model for explaining the origin of homochirality involving extraterrestrial sources of circularly polarized light (CPL) was first proposed by Rubenstein et al. (1983). The proposed scenario is as follows: (1) circularly polarized UV photons from nearby main-sequence star(s) penetrated into the natal molecular cloud that formed our solar system, (2) the photons acted on chiral molecules in the natal cloud, and resulted in an excess of one enantiomer, and (3) the excessing enantiomer was amplified until the other enantiomer had been disappeared. Since it is well-known that CPL is actually observed in the interstellar space and it has been observed in laboratories that the CPL causes an excess of one enantiomer, the scenario may work if there are chiral molecules in the interstellar molecular clouds. See Bailey (2001) for further details.

In 2016 the first report was made on the detection of a chiral molecule in space (McGuire et al. 2016). Propylene oxide was detected in the gas phase in a cold, extended molecular shell around the embedded, massive proto-stellar clusters in the Sagittarius B2 star-forming region. The fact that propylene oxide is extended suggested that similar chiral molecules may exist in a variety of astronomical sources in our Milky Way galaxy. It should be noted that chiral molecular structures of propylene oxide, S-propylene oxide and R-propylene oxide, cannot be distinguished spectroscopically.

2.4 Challenges in Searching for Amino Acids and Nucleobases in Space

2.4.1 Amino Acids

Since the first trial in detecting interstellar glycine (Brown et al. 1979), many trials to detect the simplest amino acid, glycine (NH₂CH₂COOH), were made toward Sgr B2 and other high-mass forming regions. None of them were successful. Astronomers have been searching for glycine, because if amino acids are formed in interstellar molecular clouds, significant amount of them may be delivered to planets, which would have been used as the “seed” of life on Earth and other extrasolar planets. Thus, detection of amino acids would accelerate the discussion concerning the universality of “life.”

However, the past unsuccessful searches for glycine demanded us to reconsider the search strategy. One idea to overcome this situation is to search for sources rich in precursors to glycine prior to searching for glycine. Although the chemical evolution of interstellar N-bearing complex organic molecules is not well known, methylamine (CH₃NH₂) has been proposed as a promising precursor to glycine. CH₃NH₂ can be formed from abundant molecular species, CH₄ and NH₃, on icy dust surface,

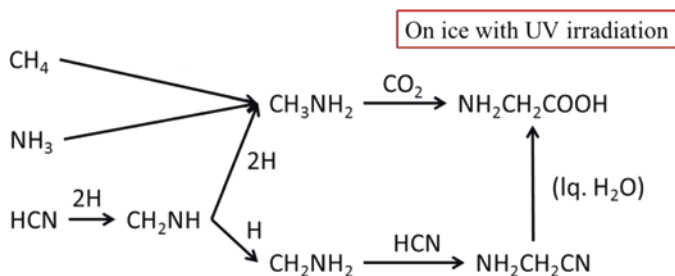


Fig. 2.3 Possible formation paths to glycine in the interstellar molecular clouds

when the dust temperature is high (Holtom et al. 2005; Kim and Kaiser 2011). Another possible route to form CH_3NH_2 is hydrogenation (addition of hydrogen) to HCN on dust surface (Kim and Kaiser 2011; Dickens et al. 1997; Theule et al. 2011): $\text{HCN} \rightarrow \text{CH}_2\text{NH} \rightarrow \text{CH}_3\text{NH}_2$. According to the laboratory study (Theule et al. 2011), the hydrogenation to HCN can form CH_3NH_2 even at the dust temperature 15 K. Figure 2.3 shows possible formation paths to CH_3NH_2 , starting from well-known abundant interstellar molecular species, HCN, CH_4 , and NH_3 . Once CH_3NH_2 is formed on the dust surface, CH_3NH_2 and another abundant molecule, CO_2 , can form glycine under UV irradiation (Kim and Kaiser 2011). Recent sensitive observations revealed sources rich in CH_2NH (Suzuki et al. 2016); some of these sources show very high abundance of CH_3NH_2 (Ohishi et al. 2018), which in turn may be suitable for searching for glycine in space.

One of the scientific purposes of ALMA is the detection of prebiotic molecules in molecular clouds where stars and planets are formed. Several projects have been going on toward the detection of glycine and other amino acids.

2.4.2 Nucleobases

A sequence of a DNA determines the corresponding sequence of amino acids forming a protein, including an enzyme; an enzyme plays an essential role to form and repair a DNA. In other words, both amino acids and nucleic acids (adenine, thymine, cytosine, guanine, and uracil) are the most important building blocks of life. Although the millimeter- and submillimeter-wave spectral lines of these nucleic acids are not known, the spectra of pyrimidine ($c\text{-C}_4\text{H}_4\text{N}_2$) have already been measured in laboratories. Pyrimidine is a six-membered ring molecule, which is known to be a direct precursor to three of the DNA and RNA bases: thymine, cytosine, and uracil.

Extraterrestrial pyrimidine was reported detected in meteoritic organic matter (Stoks and Schwartz 1981) and possibly in the carbonaceous dust of Comet Halley (Krueger et al. 1991). In the past, there were two unsuccessful searches for interstellar pyrimidine. The first search was made in 1973 by the NRAO 11 m telescope

toward Sgr B2 and Orion KL (Simon and Simon 1973); however, no upper limit to its column density was reported. The second search was made in 2003 by the James Clerk Maxwell Telescope in the 329–363 GHz range, with upper limits to its column density of a few $\times 10^{14}$ cm⁻² toward Sgr B2(N), Orion KL, and W51 (Kuan et al. 2003). More sensitive searches by, e.g., ALMA will reveal the existence of precursors to nucleic acids in space in the future.

2.5 Conclusion

In astronomy, it is well-known that organic molecules are ubiquitous in star-forming regions where planets are formed. Most of organic molecules are thought to be formed on surfaces of dust particles which will be incorporated into planets, meteorites, comets, and other small bodies. It is known that amino acids and nucleic acids are found in a few meteorites (Stoks and Schwartz 1981). The fact that the simplest amino acid, glycine, was confirmed in the coma of comet 67P/Churyumov-Gerasimenko (Altwegg et al. 2016) would make the exogenous delivery hypothesis the most plausible scenario on the origin of prebiotic organic molecules.

New and powerful radio telescope ALMA has started its full operation. The telescope is expected to reveal the existence of amino acids and other prebiotic organic molecules in star- and planet-forming regions, which are directly related to the initial compounds of chemical evolution toward the origin of life.

Acknowledgments I would like to thank Dr. Amie Elsilá for permitting the use of Fig. 2.1. This work was supported by the JSPS Kakenhi Grant Number JP15H03646. We utilized the Japanese Virtual Observatory (JVO; <http://jvo.nao.ac.jp/>) in finding relevant reference papers. This work has made use of NASA's Astrophysics Data System.

References

- Altwegg K, Balsiger H, Bar-Nun A et al (2016) Prebiotic chemicals – amino acid and phosphorus – in the coma of comet 67P/Churyumov-Gerasimenko. *Sci Adv* 2:e1600285
- Bailey J (2001) Astronomical sources of circularly polarized light and the origin of homochirality. *Orig Life Evol Biosph* 31:167–183
- Belloche A et al (2009) Increased complexity in interstellar chemistry: detection and chemical modeling of ethyl formate and n-propyl cyanide in Sagittarius B2(N). *Astron Astrophys* 499:215–232
- Belloche A, Garrod RT, Müller HSP, Menten KM (2014) Detection of a branched alkyl molecule in the interstellar medium: iso-propyl cyanide. *Science* 345:1584–1587
- Brown RD et al (1979) A search for interstellar glycine. *Mon Not R Astron Soc* 186:5P–8P
- CDMS (The Cologne Database for Molecular Spectroscopy) (2001) <http://www.astro.uni-koeln.de/cdms/molecules>. Accessed 18 Dec 2017
- Chronological Scientific Tables (Rikanenpyou) (2017) National astronomical Observatory of Japan and Maruzen

- Deamer DW, Dworkin JP, Sandford SA, Bernstein MP, Allamandola LJ (2002) The first cell membranes. *Astrobiology* 2:371–381
- Dickens JE et al (1997) Hydrogenation of interstellar molecules: a survey for methylenimine (CH_2NH). *Astrophys J* 479:307–312
- Ehrenfreund P et al (2002) Astrophysical and astrochemical insights into the origin of life. *Rep Prog Phys* 65:1427–1487
- Elsila JE, Dworkin JP, Bernstein MP, Martin MP, Sandford SA (2007) Mechanisms of amino acid formation in interstellar ice analogs. *Astrophys J* 660:911–918
- Herbst E, van Dishoeck E (2009) Complex organic interstellar molecules. *Annu Rev Astron Astrophys* 47:427–480
- Hollis JM, Lovas FJ, Jewell PR (2000) Interstellar glycolaldehyde: the first sugar. *Astrophys J* 540:L107–L110
- Hollis JM, Jewell PR, Lovas FJ, Remijan A (2004) Green bank telescope observations of interstellar glycolaldehyde: low-temperature sugar. *Astrophys J* 613:L45–L48
- Holtom PD et al (2005) A combined experimental and theoretical study on the formation of the amino acid glycine ($\text{NH}_2\text{CH}_2\text{COOH}$) and its isomer (CH_3NHCOOH) in extraterrestrial ices. *Astrophys J* 626:940–952
- Jørgensen JK et al (2012) Detection of the simplest sugar, Glycolaldehyde, in a solar-type protostar with ALMA. *Astrophys J* 757:L4 (6 pp.)
- Kaifu N et al (2004) A 8.8–50GHz complete spectral line survey toward TMC-1 I. Survey data. *Publ Astron Soc Jpn* 56:69–173
- Kim YS, Kaiser RI (2011) On the formation of amines (RNH_2) and the cyanide anion (CN^-) in electron-irradiated ammonia-hydrocarbon interstellar model ices. *Astrophys J* 729:68
- Krueger FR, Korth A, Kissel J (1991) The organic matter of Comet Halley as inferred by joint gas phase and solid phase analyses. *Space Sci Rev* 56:167
- Kuan YJ et al (2003) A search for interstellar pyrimidine. *Mon Not R Astron Soc* 345:650–656
- McGuire BA et al (2016) Discovery of the interstellar chiral molecule propylene oxide ($\text{CH}_3\text{CHCH}_2\text{O}$). *Science* 352:1449–1452
- McKellar A (1940) Evidence for the molecular origin of some hitherto unidentified interstellar lines. *Publ Astron Soc Pac* 52:187–192
- Ohishi M (2016) Search for complex organic molecules in space. *J Phys Conf Ser* 728:052002
- Ohishi M et al (2018) Detection of new methylamine (CH_3NH_2) sources: candidates for future glycine surveys. Submitted to *Publ Astr Soc Japan*
- Rubenstein E, Bonner WA, Noyes HP, Brown GS (1983) Supernovae and life. *Nature* 306:118
- Simon MN, Simon M (1973) Search for interstellar acrylonitrile, pyrimidine, and pyridine. *Astrophys J* 184:757–761
- Stoks PG, Schwartz AW (1981) Nitrogen-heterocyclic compounds in meteorites – significance and mechanisms of formation. *Geochim Cosmochim Acta* 45:563
- Suzuki T et al (2016) Survey observations of a possible glycine precursor, methanimine (CH_2NH). *Astrophys J* 825:79
- Tashiro T et al (2017) Early trace of life from 3.95 Ga sedimentary rocks in Labrador, Canada. *Nature* 549:516–518
- Theule P et al (2011) Hydrogenation of solid hydrogen cyanide HCN and methanimine CH_2NH at low temperature. *Astron Astrophys* 534:A64
- Weinreb S, Barrett AH, Meeks ML, Henry JC (1963) Radio observations of OH in the interstellar medium. *Nature* 200:829–831

Chapter 3

Chemical Interactions Among Organics, Water, and Minerals in the Early Solar System



Hikaru Yabuta

Abstract Chondritic meteorites are thought to have originated from primitive small bodies of the Solar System formed 4.5 billion years ago. Investigations on origin and chemical evolution of organic molecules in the early Solar System have been extremely improved through the chemical analyses of carbonaceous chondritic meteorites, which are derived from primitive small bodies. While carbonaceous chondrites have provided a number of significant insights on possible building blocks of life as well as the relationship between the compositions of organics and the parent body aqueous processes, precursors and locations for formation of meteoritic organics are yet to determine. It is because most of the information on the earlier stage of the Solar System history was erased by extensive degrees of aqueous alteration on the meteorite parent bodies. For constraining the origin of organic molecules in the early Solar System, it is necessary to investigate more primitive Solar System materials available to us than the typical carbonaceous chondrites, as well as it is very important to correctly determine the different evolution stages by observation to reveal the relationships between organic chemistry and mineralogy.

Through the coordinated analyses of anhydrous and hydrated Antarctic micro-meteorites (AMMs), we depicted a scenario that highly aromatic organic macromolecule (a.k.a. insoluble organic material) in carbonaceous chondrites could be a hydrolysis product of N- and/or O-rich macromolecule in a small icy body, which were formed via photochemistry in interstellar clouds or outer solar nebula. The hydrolysis probably occurred during the early stage of parent body aqueous alteration. An ultracarbonaceous AMM (UCAMM), which is highly C-rich materials and cometary origin, contains amorphous silicate grains (so-called glass with embedded metal and sulfides [GEMS]) depleted in Mg and S. The altered GEMS indicates that the cometary parent body experienced very weak aqueous alteration caused by planetesimal shock.

H. Yabuta (✉)

Department of Earth and Planetary Systems Science, Hiroshima University, Hiroshima, Japan

Department of Solar System Science, Institute of Space and Astronautical Science (ISAS),

Japan Aerospace Exploration Agency (JAXA), Sagamihara, Japan

e-mail: hyabuta@hiroshima-u.ac.jp

Keywords Organics-water-minerals interactions · Solar System · Comets · Antarctic micrometeorites · GEMS

3.1 Introduction: Enigma on Origin of Organic Molecules in the Carbonaceous Chondrites

Asteroids and comets, the remnants that did not grow into large bodies during the history of planetary formation, preserve the precursor materials in the early Solar System. These small bodies are thought to have delivered organic molecules and water to the Earth and other planets, which are necessary for planetary habitats (Chyba and Sagan 1992). Both asteroids and comets are originally derived from accretion of interstellar dusts (Fig. 3.1), which are thought to be micron-sized particles consisting of an amorphous silicate core, refractory organic inner mantle, and an outer mantle of ice (Greenberg and Li 1997). The volatile components of the dust mantle were formed by condensation of atoms and molecules in the gas phase and on dust surfaces in interstellar clouds (e.g., Tielens and Hagen 1982; Ehrenfreund and Charnley 2000; Öberg 2016). Through the subsequent photochemical reactions in interstellar dense cloud, a variety of molecules were produced (e.g., Sandford 1996; Bernstein et al. 1999; Nuevo et al. 2011). After interstellar cloud collapsed to form a young Sun, a protoplanetary disk was formed. In a protoplanetary disk, distributions of temperature and UV flux yielded the chemical diversity of the dusts, followed by dust growth to form planetesimals (Nomura et al. 2007; Ciesla and Sandford, 2012). As a result, rocky planetesimals, so-called asteroids, were formed in inner Solar System, whereas icy planetesimals, so-called comets, were formed in outer Solar System, such as Kuiper Belt and Oort cloud (Fig. 3.1). Subsequently, asteroids experienced internal heating due to the decay of short-lived radiogenic nuclides (e.g., ^{26}Al). This event caused aqueous alteration (0–150 °C) (e.g., Brearley 2006, and references therein) and thermal metamorphism (200–700 °C) (e.g., Huss et al. 2006, and references therein) in the asteroid parent bodies, which lasted for several million years (e.g., Fujiya et al. 2013). Asteroids also experienced shorter duration thermal metamorphism induced by impacts. Both secondary processes and radial distance resulted in chemical compositions of asteroids, which are reflected by the reflectance spectral types of asteroids: stony objects (S-type), dark carbonaceous objects (C-type), very dark objects without spectral feature of phyllosilicates (D-type), metallic objects (M-type), and so on (Tholen and Barucci 1989) (Fig. 3.1). Comets could experience these secondary processes as well, although they were less altered due to the low inner temperature of icy bodies. Conditions, timescales, and durations of the secondary processes depend on the sizes, structure, and heat sources of parent bodies, which increased chemical diversity in asteroids and comets. Thus, the Solar System can be said to be “a chemical factory” of life’s building blocks.

To date, primitive carbonaceous chondrite meteorites have been frequently investigated for understanding origin and evolution of organic molecules in the early Solar System. In general, carbonaceous chondrite contains several weight% total organic carbon, which can be divided into soluble (free) and acid-insoluble

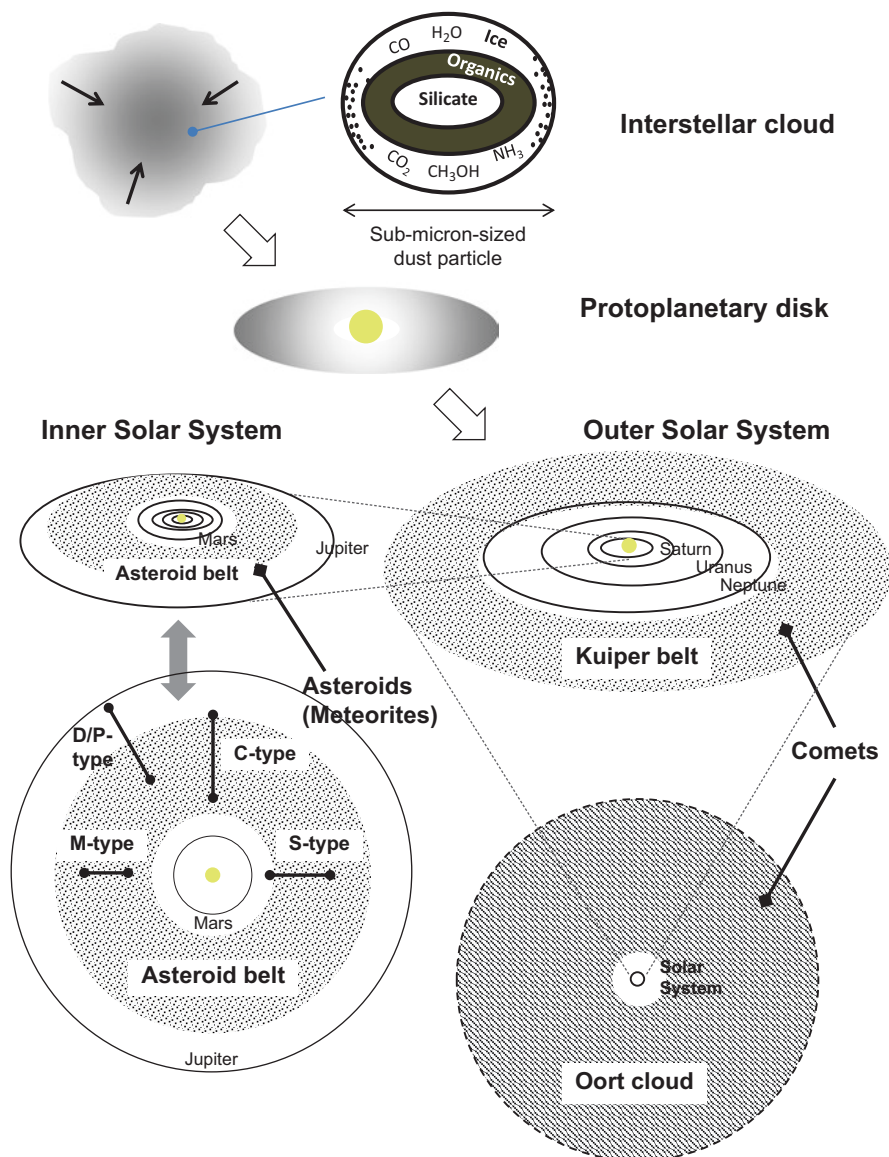


Fig. 3.1 Outline of Solar System evolution from interstellar clouds to asteroids and comets

(refractory) fractions. The insoluble organic matter (IOM) accounts for more than a half of total organic carbon. Although the intact molecular structure of IOM is still unknown due to its complex, macromolecular configuration, IOM is composed of aromatic molecular network crosslinking with short-branched aliphatic chains and various oxygen functional groups (Cody et al. 2002; Glavin et al. 2018) (Fig. 3.2). Soluble organic molecules have been historically very well studied represented by

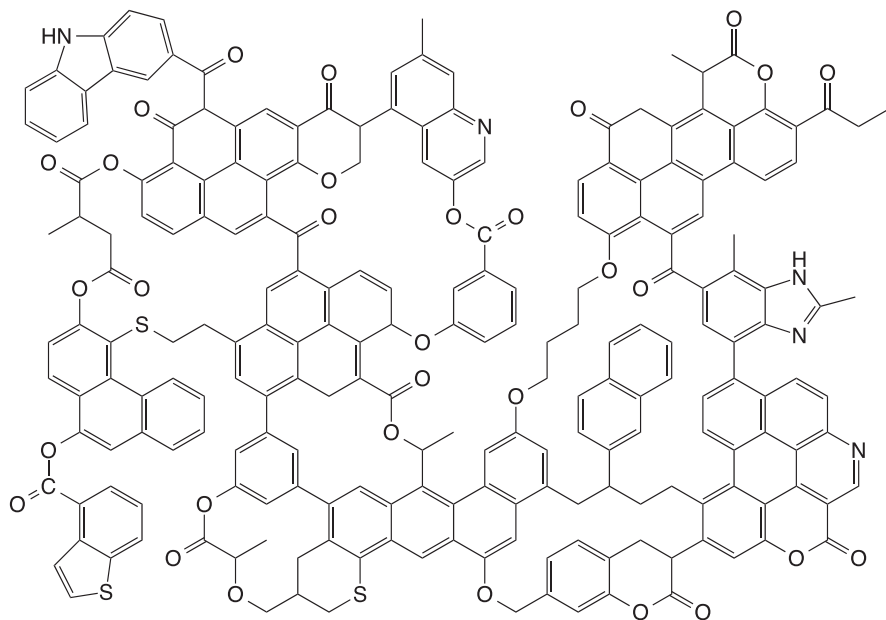


Fig. 3.2 A model of the molecular structure of the IOM in primitive carbonaceous chondrites (i.e., CM group) that is consistent with what is known about its functional group chemistry (Glavin et al. 2018)

amino acids and a variety of compounds of biochemical interests (Pizzarello et al. 2006; Glavin et al. 2018). The total abundances of the individual molecules are very low (from ppb to ppm levels), and the sum abundance of the molecules found does not exceed the half of total organic carbon, implying that there still exist unidentified molecules (Schmitt-Kopplin et al. 2010).

Through 50 years of the studies on meteoritic organics, a number of hypotheses have suggested for origin of refractory organic macromolecules in carbonaceous chondrites. Alexander et al. (2007) suggested that IOM is derived from the refractory organics formed via photochemistry in the extreme cold environments, such as interstellar cloud or outer solar nebula, based on the observed enrichments of deuterium (D) and ^{15}N in IOM in meteorites and interplanetary dust particles (IDPs). On the other hand, the potential roles of Fischer-Tropsch (FT) synthesis (e.g., Hill and Nuth 2003) and irradiation reaction of nebular gas (i.e., CO, N_2 , H_2O) (Kuga et al. 2015) for the formation of meteoritic IOM have been argued. It was recently suggested that the first IOM might have been formed via hydrothermal reaction of formaldehyde and ammonia on the meteorite parent body based on the spectroscopic similarity between the experimentally synthesized products and meteoritic IOM (Cody et al. 2011; Kebukawa et al. 2013). Thus, there are multiple possible scenarios proposed, and the consensus has yet to be reached. Since there may be diversity in precursor molecules and locations of organic materials, it is very important to determine the specific sources for specific molecules found in individual meteorites and other extraterrestrial samples.

3.2 A Missing Link in the Early Solar System: Chemical Evolution of Organic Molecules from Solar Nebula to Planetesimals

Most of carbonaceous chondrites experienced extensive aqueous alteration on their parent bodies, and thus it is hard to find characteristics of precursor organic materials from the meteorites. It is therefore necessary to investigate more primitive Solar System materials than the typical carbonaceous chondrites, to constrain the origin of organic molecules in the early Solar System. Those pristine samples include comets and comet-derived materials, such as cometary dusts, stratospheric interplanetary dust particles (IDPs), Antarctic micrometeorites (AMMs), etc., since comet is an icy body formed at very low temperature that preserves the materials in the interstellar cloud and solar nebula under better condition (Yabuta et al. 2018). Comet dust sample return mission, Stardust, detected O- and N-rich refractory organic macromolecules, glycine and amines, from the comet Wild 2 dust particles for the first time (e.g., Sandford et al. 2006; Cody et al. 2008; Glavin et al. 2008). The mission to rendezvous with Jupiter-family comet, Rosetta, for the first time discovered dark refractory organic materials on the surface of 67P/Churyumov-Gerasimenko (e.g., Quirico et al. 2016). These results supported the relationship between comets and IDPs. In particular, chondritic porous (CP)-type IDPs have been regarded as short-period comet origin, based on their fragile structure, higher amounts of carbon and volatiles than those in meteorites, and amorphous silicate so-called glass with embedded metal and sulfides (GEMS) (Keller and Messenger 2011). The IDPs show a wide range of hydrogen and nitrogen isotopic compositions (Messenger 2000).

Another clue is the chemical relationship between organics and minerals. For example, the presence of GEMS is a good indicator of the stage prior to aqueous activities on the parent body, since the amorphous silicate is rapidly changed to hydrated silicate minerals when it is exposed to liquid water (Nakamura-Messenger et al. 2011). Thus, the organic materials associated with GEMS could be the precursor materials in interstellar or solar nebula (Keller and Messenger 2011). In contrast, hydrated silicate mineral is a robust indicator of aqueous process in the parent body, and the organics associated with those minerals could be a secondary alteration product. Those classifications based on mineralogy would enable correct understanding of chemical evolution of organic molecules during the process from solar nebula to planetesimals.

This chapter will address the precursors of organic materials and minerals in planetesimals, through the reviews of our recent investigations using the AMMs collected from the surface snow near the Dome Fuji Station, Antarctica. The AMMs include the samples of cometary origin, as Noguchi et al. (2015) discovered chondritic porous AMMs containing GEMS, enstatite whiskers, and organic nanoglobules, which were characteristic to CP IDPs collected at the stratosphere (Fig. 3.3). We carried out coordinated comprehensive chemical analyses of the AMMs, for maximizing the data of organic chemistry, mineralogy, and isotope cosmochemistry from a single AMM particle.

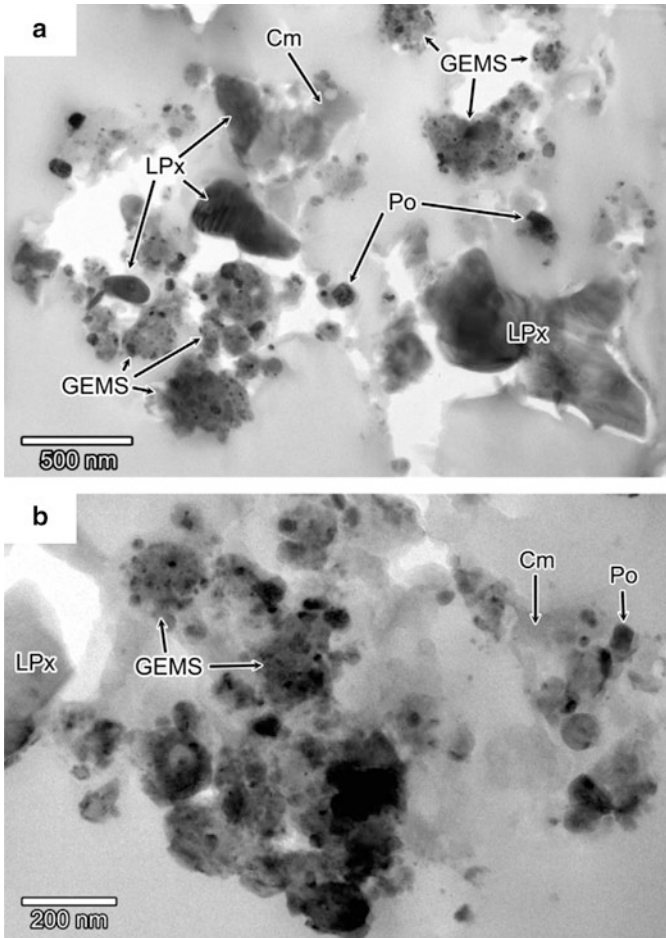


Fig. 3.3 Bright-field image of an ultrathin section of (a) a present CP MM D051B13 and (b) a CP IDP L2021. The CP MM contains abundant GEMS. Low-Ca pyroxene (LPx) and pyrrhotite (Po) (Noguchi et al. 2015)

3.3 Different Stages of Parent Body Aqueous Alteration Recorded in Antarctic Micrometeorites

Figure 3.4 shows transmission electron microscopy (TEM) images of the ultrathin sections of two AMMs. AMMs containing GEMS (Fig. 3.4a) were classified as anhydrous micrometeorites (MMs). The organic materials ($\sim 10 \times 3 \mu\text{m}$) account for a major part of the anhydrous MM, and they are associated with olivine and pyrrhotite. On the other hand, AMMs containing phyllosilicate (clay minerals) were classified as hydrous MMs (Fig. 3.4b). The amounts of organic materials are small compared to those in the anhydrous MM, and they are associated with phyllosilicate

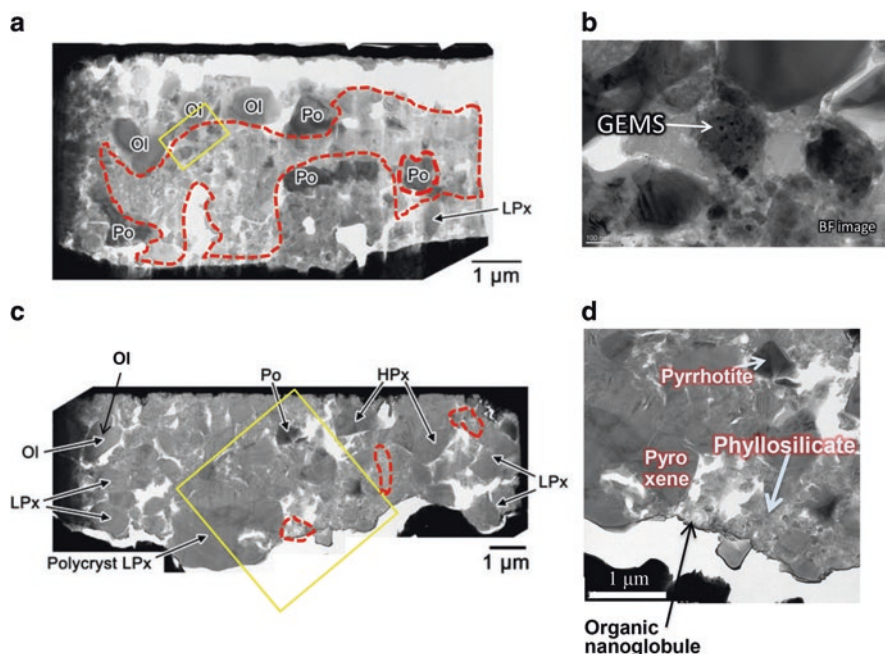


Fig. 3.4 TEM images of ultrathin sections of the AMMs obtained by focused ion beam (FIB). Coarse-grained minerals and aggregates of minerals are labeled. (a) D10IB009, (b) an enlarged image of yellow square region of (a, c) D10IB163, (d) an enlarged image of yellow square region of (c) (Noguchi et al. 2017). Abbreviations: *Ol* olivine, *LPx* low-Ca pyroxene, *HPx* high-Ca pyroxene, *GEMS* glass embedded metal and sulfide, *Po* pyrrhotite. The organic materials are surrounded by red broken lines

and pyroxene. This difference probably indicates that the precursor organic materials, which were present before planetesimal formation, were depleted and modified by parent body aqueous alteration.

X-ray absorption near-edge structure (XANES) is a method to assess the types of the organic functional groups constituting a macromolecular sample (Cody et al. 2008). XANES spectra are obtained by scanning x-ray energy and measuring the absorbed x-ray intensity specific to the electronic structures (chemical bonds) of the atom that absorbed the x-ray (Stöhr 1991). The absorbed x-ray intensity corresponds to photoexcitation of carbon 1s electrons to unoccupied electronic state. Combination with a scanning transmission x-ray microscope (STXM) with high spatial resolution (<30 nm) enables the measurement of XANES spectra of submicron-sized regions of organic macromolecules. Application of STXM-XANES to AMMs revealed that organic material in the anhydrous MM is enriched in N- and O-bearing functional groups, such as nitrile ($-\text{C}\equiv\text{N}$) or purine-pyrimidine, carboxyl groups ($-\text{COOH}$), and aliphatic carbon ($-\text{CH}_x$) (Fig. 3.5). This composition was prominent particularly in the regions indicated with white arrows. According to TEM observation, some of the regions were organic nanoglobules.

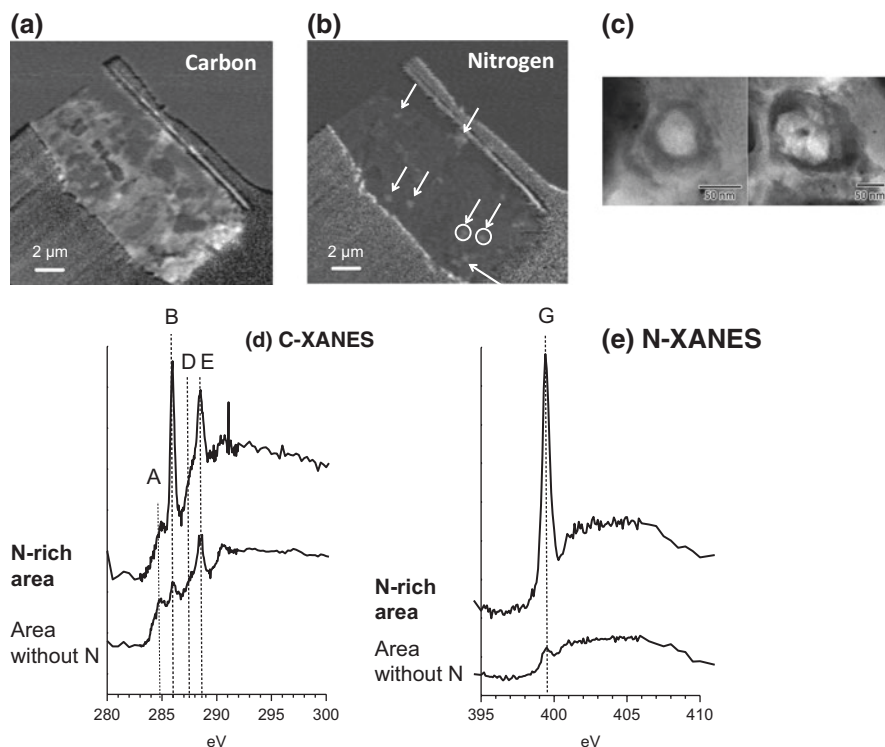


Fig. 3.5 (a) Carbon- and (b) nitrogen-elemental maps of anhydrous AMM (D10IB009) acquired by a scanning transmission x-ray microscope (STXM) (optical density images). Nitrogen-enriched areas are indicated by white arrows. Circle regions correspond to (c) bright-field images of organic nanoglobules. (d) Carbon- and (e) nitrogen-K edge XANES spectra of a nitrogen-enriched area and an area without enrichment of nitrogen in D10IB009 (Noguchi et al. 2017). Peak A, $1s\pi^*$ transition of aromatic carbon ($C=C^*$) at 285 eV; peak B, $1s\pi^*$ transition of N-heterocycles ($C-N^*=C$) and/or nitrile ($C\equiv N^*$) at 286 eV; peak D, $1s3p/s^*$ transition of aliphatic carbon at CH_x-C at ~ 287.5 eV; peak E, $1s\pi^*$ transition of carbonyl carbon in carboxyl or ester ($OR(C^*=O)C$) at ~ 288.4 eV; peak G, $1s\pi^*$ transition of N-heterocycles ($C-N^*=C$) and/or nitrile ($C\equiv N^*$) at 399.4 eV

From this anhydrous MM, enrichments of D ($\delta D = +8000 \sim +10,000\text{‰}$) and ^{15}N ($\delta^{15}N = +600 \sim +1000\text{‰}$) were detected by secondary ion mass spectrometry (SIMS) (Fig. 3.6). These values were comparable with the pristine IDPs (e.g., Messenger 2000; Busemann et al. 2009) and comet Wild 2 dust particles (e.g., Mckeegan et al. 2006; De Gregorio et al. 2010). It has been suggested that the isotopic anomalies of H in the small body organic materials are derived from molecular deuterium enrichment caused by ion-molecule reactions under very low temperature (10–50 K) in interstellar molecular clouds (e.g., Millar et al. 1989) and in protoplanetary disks (e.g., Aikawa and Herbst, 1999). Although the sources of ^{15}N enrichment in the small body organic materials are clearly determined, contribution of interstellar chemistry is indicated (e.g., Charnley and Rodgers 2002; Aleon and Robert 2004). Thus, it is very likely that organic material from the anhydrous MM

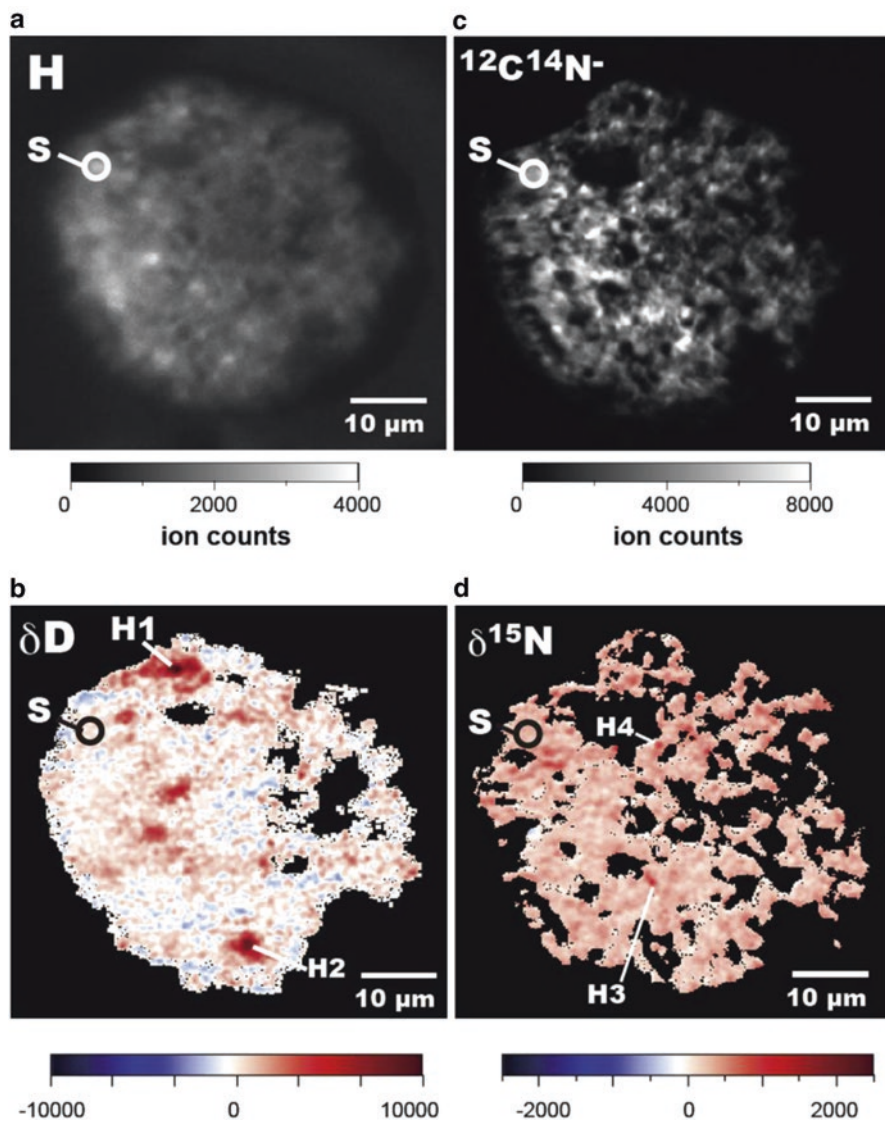


Fig. 3.6 Isotopographs of anhydrous AMM (D10IB009). (a) H. (b) δD_{SMOW} . H1 and H2 denote δD hotspots (H1, 10,000‰, and H2, 8000‰). (c) $^{12}C^{14}N^-$ and (d) $\delta^{15}N_{air}$. H3 and H4 denote $\delta^{15}N$ hotspots (H3, 1000‰, and H4, 600‰). Circles (S) of each image (a–d) denote the spot analysis area by SIMS to estimate the instrumental mass fractionation (Noguchi et al. 2017)

is originated from interstellar cloud or outer solar nebula. In particular, their isotopic values were comparable with δD of cometary HCN (Meyer et al. 1998) and $\delta^{15}N$ of cometary CN (Schultz et al. 2008), and thus the N- and O-rich functional group chemistry of the anhydrous MM may have been derived from photochemistry of

HCN- or CN-ice. D- and ^{15}N -rich organic nanoglobules may have been also formed during the UV irradiation of ice grains under the cold environment, as suggested by Nakamura-Messenger et al. (2006).

Comparing the C- and N-XANES spectra among the anhydrous and hydrous AMMs and insoluble organic matter (IOM) in Murchison carbonaceous chondritic meteorite, all the samples showed typical three peaks of aromatic carbon, aromatic ketone, and carboxyls with a shoulder of aliphatic carbon (Fig. 3.7). However, the relative peak intensities were slightly different between anhydrous and hydrous MMs, i.e., the peaks of carboxyl groups are higher in the two anhydrous MMs compared to those of hydrous MMs. On the other hand, the spectra of the hydrous MMs were enriched in aromatic carbon, which is similar to those of Murchison IOM. Exceptionally, the spectrum of anhydrous AMM-D10IB004 is rather similar to those of hydrous AMMs.

Characteristic features of organic materials and minerals in the AMMs are summarized in Table 3.1. All the three anhydrous AMMs (D10IB009, D10IB356, D10IB004) characterized by GEMS, Fe-Ni metal, and sulfide are very likely cometary origin having no record of aqueous alteration. Based on the presence of amorphous silicate and minor Fe-rich serpentine, hydrated AMM (D10IB178) is derived from carbonaceous CR3 chondrite-type parent body that experienced weak aqueous

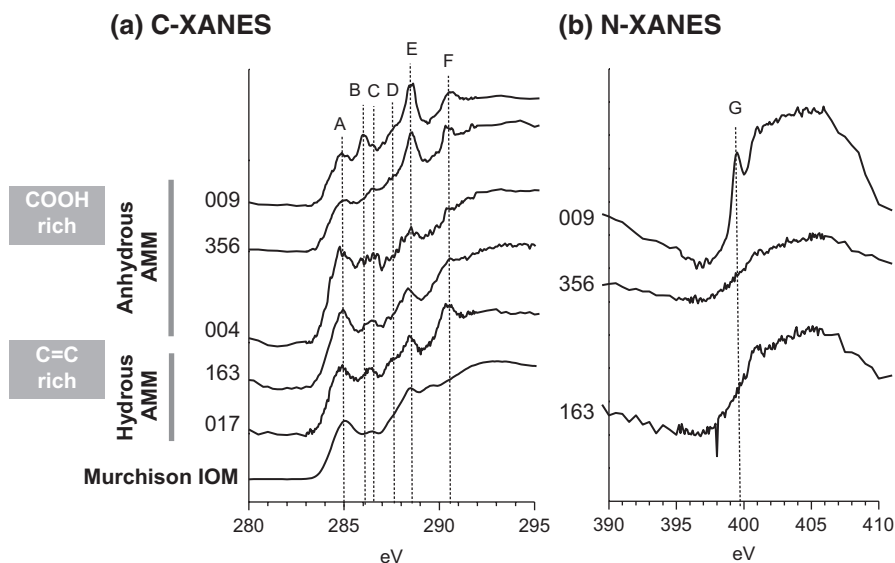
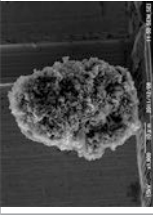
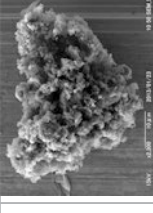
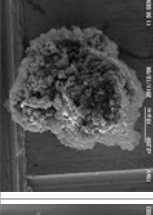
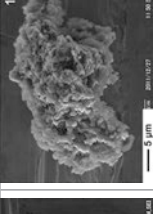
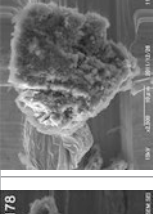
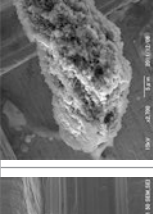


Fig. 3.7 Averaged carbon K edge – XANES spectra of five AMMs (D10IB009, 356, 004, 163, and 017) and Murchison meteorite IOM (Noguchi et al. 2017). Peak A, $1s\pi^*$ transition of aromatic carbon ($\text{C}=\text{C}^*$) at 285 eV; peak B, $1s\pi^*$ transition of N-heterocycles ($\text{C}-\text{N}^*=\text{C}$) and/or nitrile ($\text{C}\equiv\text{N}^*$) at ~ 286.6 eV; peak C, $1s\pi^*$ transition of vinyl-keto carbon ($\text{C}=\text{C}-\text{C}^*=\text{O}$) at ~ 286.6 eV; peak D, $1s3p/s^*$ transition of aliphatic carbon at CH_x-C at ~ 287.5 eV; peak E, $1s\pi^*$ transition of carbonyl carbon in carboxyl or ester ($\text{OR}(\text{C}^*=\text{O})\text{C}$) at ~ 288.4 eV; peak F, $1s\pi^*$ transition of carbonate or carbonic acid ($\text{RO}(\text{C}=\text{O})\text{OR}'$) at 290.6 eV

Table 3.1 Characteristic features of organic materials and minerals in AMMs investigated in Noguchi et al. (2017)

Sample ID	CP AMMs (anhydrous)			Fluffy fine grained AMMs (hydrous)		
	D10IB009	D10IB356	D10IB004	D10IB178	D10IB163	D10IB017
<i>Mineralogy</i>	 GEMS, Fe-Ni metal, Sulfide, Olivine, low-Ca pyroxene	 GEMS, Fe-Ni metal, Sulfide, Olivine, low-Ca pyroxene	 GEMS, Fe-Ni metal, Sulfide, Olivine, low-Ca pyroxene	 Amorphous silicate, Olivine, low-Ca pyroxene, Fe-rich saponite, minor Fe-rich serpentine	 Olivine, low-Ca pyroxene, Fe-rich saponite, minor Fe-rich serpentine	 Olivine, low-Ca pyroxene, Fe-rich saponite, minor Mg-rich serpentine, Magnesite
<i>Organic chemistry</i>	Carboxyls (COOH), Aliphatic, Nitrile (CN) or N-heterocycles	COOH, Aliphatic	Aromatic, COOH Aromatic ketone, COOH Chondritic IOM-like	–	Aromatic, COOH, Aromatic ketone, Chondritic IOM-like	Aromatic, COOH, Aromatic ketone, Chondritic IOM-like
<i>Isotope</i>	Abundant globules $\delta^{15}\text{N} = \sim 600\text{-}1000\text{‰}$ $\delta\text{D} = \sim 8000\text{-}10000\text{‰}$	–	$\delta^{15}\text{N} = \sim 300\text{‰}$ $\delta\text{D} = \text{normal}$	–	One globule –	–
<i>Aqueous alteration</i>	No	No	No	Weak	Weak	Moderate

alteration. The other hydrated AMMs (D10IB163 and D10IB017) contain abundant coarse-grained anhydrous minerals with lesser amounts of hydrate silicate minerals which also experienced weak to moderate aqueous alteration. Assuming that those AMMs are reflected by contiguous evolution stages, it is suggested that organic materials containing aliphatic carbon, COOH, and/or N-heterocycles in the anhydrous AMMs reflect the precursor molecule that were formed in interstellar cloud or outer solar nebula. The O- and N-rich molecular compositions could have been hydrolyzed by the subsequent parent body aqueous alteration and converted to highly aromatic, meteoritic organics like compositions as seen in hydrated AMMs. The anhydrous MM (D10IB004) may have experienced a very weak degree of aqueous process which did not affect silicate but sufficiently promoted modification of organic materials. The processes could occur in very primitive icy asteroids that bear comet-like feature, e.g., D-type asteroids.

3.4 Ultracarbonaceous Antarctic Micrometeorites: New Type of Cometary Material?

Ultracarbonaceous Antarctic micrometeorites (UCAMMs) are unique extraterrestrial materials that contain a large amount of carbonaceous materials (Fig. 3.8). These MMs were for the first time discovered by Nakamura et al. (2005). One of the UCAMMs contains light noble gases with a solar wind origin, and two are abundant in presolar grains (Yada et al. 2008; Floss et al. 2012). Other UCAMMs have been independently found from the Concordia Station, Antarctica, by the French-Italian team. It was suggested that organic materials in the UCAMMs could be formed in the outer solar nebula (Duprat et al. 2010), based on the identification of crystalline silicates which were thought to be derived from solar origin and are associated with D-rich organic materials. Dartois et al. (2013) reported D- and ^{15}N -rich UCAMMs

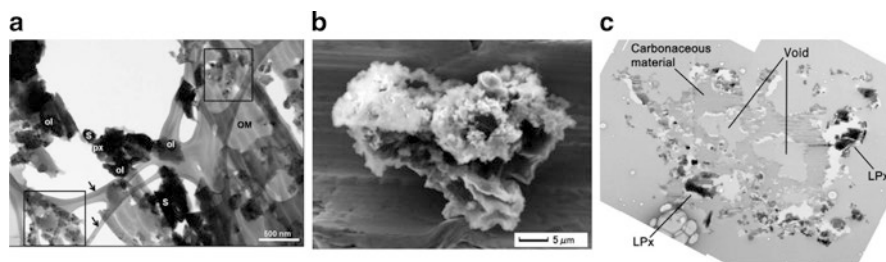


Fig. 3.8 (a) Bright-field (BF) TEM image of a UCAMM collected from Concordia Station. The organic materials are indicated as black arrows (Duprat et al. 2010). (b) Secondary electron image of the UCAMM (D05IB80) placed on a platinum plate (Yabuta et al. 2017), (c) BF-TEM images of ultramicrotomed sections of the UCAMM D05IB80 (Yabuta et al. 2017). *Abbreviations:* ol Mg-rich crystalline olivine, *px* Mg-rich crystalline pyroxene, *S* Fe-Ni sulfides, *OM* organic material, *LPx* low-Ca pyroxene. GEMS-like objects are surrounded by black squares

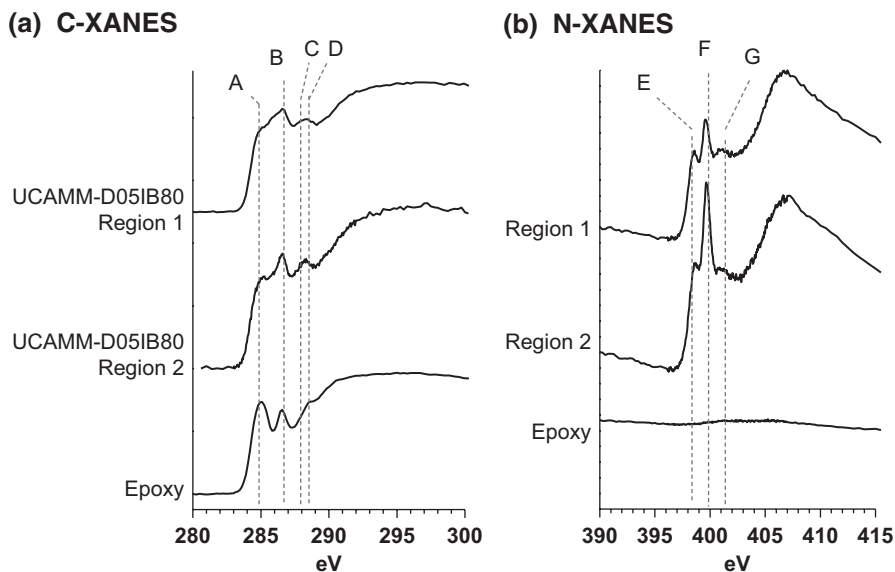


Fig. 3.9 (a) C- and (b) N-XANES spectra of the regions 1 and 2 of the UCAMM D05IB80 and epoxy. Peak A, $1s\pi^*$ transition for aromatic carbon ($C=C^*$) at 285.1 eV; peak B, $1s\pi^*$ transition for N-heterocycles ($C-N^*=C$), nitrile ($C\equiv N^*$), or vinyl-keto carbon ($C=C-C^*=O$) at ~ 286.6 eV; peak C, $1s3p/s^*$ transition for aliphatic carbon at CH_x-C at ~ 287.5 eV; peak D, $1s\pi^*$ transition for carbonyl carbon in amide ($NH_x(C^*=O)C$) at ~ 288.0 – 288.2 eV and/or $1s\pi^*$ transition for carbonyl carbon in carboxyl or ester ($OR(C^*=O)C$) at ~ 288.4 – 288.7 eV; peak E, $1s\pi^*$ transition for imine ($C=N^*$) at 398.8 eV; peak F, $1s\pi^*$ transition for N-heterocycles ($C-N^*=C$) and/or nitrile ($C\equiv N^*$) at ~ 399.7 eV; and peak G, $1s\pi^*$ transition for amide ($N^*H_x(C=O)C$) or $1s3p/s^*$ transition for amino ($C-N^*H_x$) at 401.5 eV

and suggested organic material in UCAMMs was formed in the Oort cloud by irradiation of ice rich in CH_4 and N_2 .

Most of the UCAMMs have common chemical features that are regarded as cometary origin (Duprat et al. 2010; Dartois et al. 2013; Yabuta et al. 2017). High abundances of GEMS and absence of hydrated minerals are the mineralogical characteristics of UCAMMs, which are clearly distinct from most of primitive carbonaceous chondrites. Organic materials in UCAMMs are larger than 100 times those in primitive carbonaceous chondrites, and they contain a variety of nitrogen-bearing functional groups such as nitrile, aromatic N, amide, and imine (Fig. 3.9) (Yabuta et al. 2017). The N/C ratios of organic materials in UCAMMs are five times higher than those in primitive carbonaceous chondrites (Dartois et al., 2013; Yabuta et al. 2017). As is the case of IDPs, UCAMMs show a wide range of H and N isotopic compositions, from values extremely rich in heavy isotopes (Duprat et al. 2010; Dartois et al. 2013) to normal values with terrestrial levels (e.g., Yabuta et al. 2017).

The N/C and O/C ratios of organic material from UCAMM DO05IB80 are very similar to those of comet Wild 2 dust particles (Fig. 3.10) (Yabuta et al. 2017). On the other hand, the values are very different from those of the UV irradiation

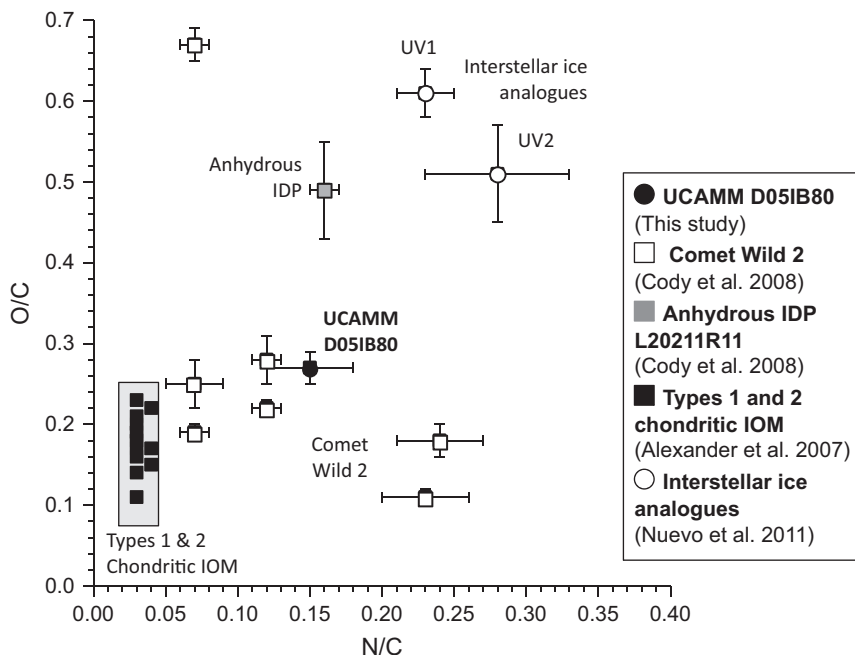


Fig. 3.10 N/C versus O/C ratios of organics in the UCAMM D05IB80 (●, this study), the comet Wild 2 dust particles (□, Cody et al. 2008), the anhydrous IDP L20211R11 (■, Cody et al. 2008), type 1 and 2 chondritic insoluble organic solids (■, Alexander et al. 2007), and the UV irradiation products from interstellar ice analogues (O, Nuevo et al. 2011) (UV1 $\text{H}_2\text{O}:\text{CH}_3\text{OH}:\text{CO}:\text{NH}_3 = 100:50:1:1$, UV2 $\text{H}_2\text{O}:\text{CH}_3\text{OH}:\text{CO}:\text{NH}_3:\text{C}_3\text{H}_8 = 100:50:1:1:10$). The ratios were estimated from the fitting of C-, N-, and O-XANES spectra (Yabuta et al. 2017)

products from interstellar ice analogues. Thus, interstellar photochemistry alone would not be the process responsible for the formation of the UCAMM. Yabuta et al. (2017) suggested that a very small amount of fluid on a cometary nucleus or icy asteroid must have been necessary for the formation of the UCAMM, based on the following multiple evidences: (i) the presence of sulfur in an entire region of organic materials in UCAMM (organic sulfurization), (ii) deformation and aggregation of organic nanoglobules (change in osmotic pressure) (Fig. 3.10), (iii) inorganic thin layers at the surface of organic materials in UCAMM (precipitation of dissolved ions) (Fig. 3.10), and (iv) depletion of Mg and S from GEMS (Fig. 3.11).

Possible heat sources for the generation of liquid water in icy small bodies include (i) short-lived radioactive nuclides, (ii) perihelion passage (Nakamura-Messenger et al. 2011), (iii) collisions of planetesimals (Cody et al. 2011), and (iv) reduction of the freezing point by the presence of solutes, e.g., ammonia (Pizzarello et al. 2011) or methanol. The condition of aqueous alteration of the UCAMM can be estimated by referring to the hydrothermal experiment of anhydrous IDPs by Nakamura-Messenger et al. (2011). In the experiment, the alteration of amorphous

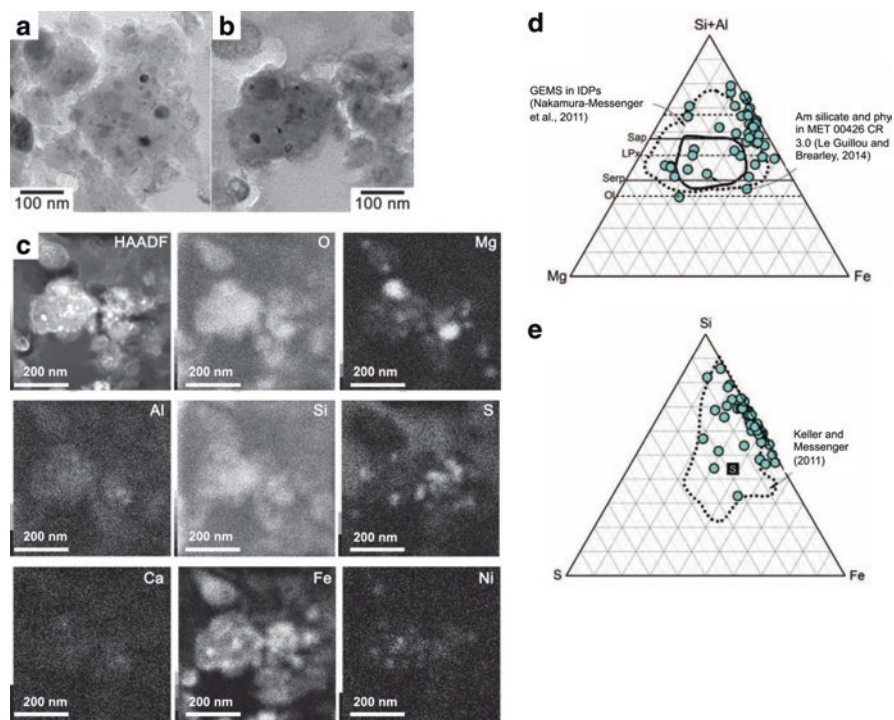


Fig. 3.11 (a, b) BF-TEM images of GEMS grains in the UCAMM D05IB80. (c) HAADF-STEM image and elemental distribution maps of the same GEMS grains in (b). In comparison with HAADF-STEM image and O, Si, and Fe maps, depletion of Mg and S are observed. (d) [Si + Al]-Mg-Fe ternary diagram and (e) Si-S-Fe ternary diagram of GEMS grains in the UCAMM D05IB80 (Yabuta et al. 2017)

silicate into hydrous phyllosilicate proceeded extremely quickly, at 25–160 °C in 12–24 h under basic pH conditions (pH = 12). Since hydrous silicates are not identified in UCAMM D05IB80, the UCAMM could have experienced a shorter duration reaction at lower temperature, lower pH, and/or slightly oxidizing conditions compared to the experiments. Considering that the degree of alteration would have been much lower than the aqueous alteration on the typical carbonaceous chondritic meteorite parent bodies (i.e., CM and CI groups), which lasted for several million years (e.g., Fujiya et al. 2013), planetesimal collisions are most likely cause to produce a very weak degree of aqueous alteration in a short duration. Weak alteration in short duration was probably caused by planetesimal shock that locally melted cometary ice grains and released water that dissolved the organics; the fluid would have not been mobilized because of the very low thermal conductivity of the porous icy body. This event allowed the formation of the large organic puddle of the UCAMM, as well as organic matter sulfurization, formation of thin membrane-like layers of minerals, and deformation of organic nanoglobules.

3.5 Conclusion

Comprehensive investigations on organic chemistry and mineralogy of Antarctic micrometeorites have enabled determination of the history from solar nebula to planetesimals, including even early stage where phyllosilicates are not affected but organics are sufficiently changed. This submicron-to-nanoscale observation of the organic material-water-mineral interactions in the small body materials would be expected to find an evidence for the transition from comets to asteroids, which may have been caused by Jupiter's migration in the history of Solar System formation (Walsh et al. 2012). Further investigations of greater numbers of primitive meteorites, micrometeorites, and IDPs would be required for understanding the statistical features of comet-asteroid continuum.

In addition, combination of the small body explorations with laboratory experiments would provide clear insights into the origin and evolution of organic molecules in the Solar System. A Japanese asteroid sample return mission, Hayabusa2, aims to reveal the origins and evolution of the Earth-life and ocean, by investigating a C-type asteroid. The spacecraft has arrived at the asteroid (162173) Ryugu on June 27, 2018, will collect the surface of the asteroid during its 18-month stay, and will return the sample to the Earth in the end of 2020 (Tachibana et al. 2014; Yabuta 2018). National Aeronautics and Space Administration (NASA)'s asteroid sample return mission, OSIRIS-REx, will target the B-type asteroid (that is fallen into C-type in a broad sense) (101955) Bennu, which is also thought to be enriched in carbon and water, and return the collected sample to the Earth in 2023 (Lauretta 2017).

Regarding the missions related to cosmic dust science, laboratory analyses of the particles collected at the low Earth orbit by using silica aerogels have been ongoing, as a part of Japanese astrobiology experiment on the International Space Station (Tanpopo mission) (Yano et al. 2017). Demonstration and Experiment of Space Technology for INterplanetary voYage Phaethon fLYby dUST science, DESTINY+ (DESTINY-plus), is a Japanese mission to flyby of Geminids parent (3200) Phaethon and to conduct in situ dust analyses that is proposed to be launched in 2022 (Arai et al. 2018). The spacecraft will be equipped with Dust Analyzer (impact ionization time-of-flight mass spectrometry (ToF-MS)) as a payload to measure the velocity, orbital, sizes, and masses and chemical composition of interplanetary and interstellar dust particles during deep space cruising phase and the dusts from Phaethon. NASA's Comet Astrobiology Exploration Sample Return (CAESAR) mission is planned to be launched in 2024, collects surface material from the nucleus of comet 67P/Churyumov-Gerasimenko, and returns the sample to the Earth in 2038 (Squyres et al. 2018). The continuous, systematic explorations of small bodies in different evolution stages will elucidate the substantial role and mechanism of exogenous delivery of organic molecules and water to the early Earth.

References

- Aikawa Y, Herbst E (1999) Deuterium fractionation in protoplanetary disks. *Astrophys J* 526:314–326
- Aléon J, Robert F (2004) Interstellar chemistry recorded by nitrogen isotopes in solar system organic matter. *Icarus* 167:424–430
- Alexander CMO'D, Fogel M, Yabuta H, Cody GD (2007) The origin and evolution of chondrites recorded in the elemental and isotopic compositions of their macromolecular organic matter. *Geochim Cosmochim Acta* 71:4380–4403
- Arai T, Kobayashi M, Ishibashi K, Yoshida F, Kimura H, Wada K, Senshu H, Yamada M, Okudaira O, Okamoto T, Kameda S, Srama R, Krüger H, Ishiguro M, Yabuta H, Nakamura T, Watanabe J, Ito T, Ohtsuka K, Tachibana S, Mikouchi T, Komatsu M, Nakamura-Messenger K, Sasaki S, Hiroi T, Abe S, Urakawa S, Hirata N, Demura H, Komatsu G, Noguchi T, Sekiguchi T, Inamori T, Yano H, Yoshikawa M, Ohtsubo T, Okada T, Iwata T, Nishiyama K, Toyota H, Kawakatsu Y, Takashima T (2018) DESTINY+ mission: Flyby of Geminids parent asteroid (3200) Phaethon and in-situ analyses of dust accreting on the Earth. The 49th Lunar and Planetary Science Conference, Abstract #2570
- Bernstein MP, Sandford SA, Allamandola LJ, Gillette JS, Clemett SJ, Zare RN (1999) UV irradiation of polycyclic aromatic hydrocarbons in ices: production of alcohols, quinones, and ethers. *Science* 283:1135–1138
- Brearley A. J. (2006) The action of water. Meteorites and the early solar system II (Lauretta D. S. McSween Jr. H.Y 587–624 Tucson: University of Arizona Press
- Busemann H, Nguyen AN, Cody GD, Hoppe P, Kilcoyne ALD, Stroud RM, Zega TJ, Nittler LR (2009) Ultra-primitive interplanetary dust particles from the comet 26P/Grigg-Skjellerup dust stream collection. *Earth Planet Sci Lett* 288:44–57
- Charnley SB, Rodgers SD (2002) The end of interstellar chemistry as the origin of nitrogen in comets and meteorites. *Astrophys J* 569:L133–L137
- Chyba C, Sagan C (1992) Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature* 355:125–132
- Ciesla F, Sandford SA (2012) Organic synthesis via irradiation and warming of ice grains in the solar nebula. *Science* 336:452–454
- Cody GD, Alexander CMO'D, Tera F (2002) Solid-state (^1H and ^{13}C) nuclear magnetic resonance spectroscopy of insoluble organic residue in the Murchison meteorite: a self-consistent quantitative analysis. *Geochim Cosmochim Acta* 66:1851–1865
- Cody GD, Ade H, Alexander CMO'D, Araki T, Butterworth A, Fleckenstein H, Flynn G, Gilles MK, Jacobsen C, Kilcoyne ALD, Messenger K, Sandford SA, Tylliszczak T, Westphal AJ, Wirick S, Yabuta H (2008) Quantitative organic and light-element analysis of comet 81P/Wild 2 particles using C-, N-, and O- XANES. *Meteor Planet Sci* 43:353–365
- Cody GD, Heying E, Alexander CMOD, Nittler LR, Kilcoyne ALD, Sandford SA, Stroud RM (2011) Establishing a molecular relationship between chondritic and cometary organic solids. *Proc Natl Acad Sci U S A* 108:19171–19176
- De Gregorio BT, Stroud RM, Nittler LR, Alexander CMO'D, Kilcoyne ALD, Zega TJ (2010) Isotopic anomalies in organic nanoglobules from comet 81P/wild 2: comparison to Murchison nanoglobules and isotopic anomalies induced in terrestrial organics by electron irradiation. *Geochim Cosmochim Acta* 74:4454–4470
- Dartois E, Engrand C, Brunetto R, Duprat J, Pino T, Quirico E, Remusat L, Bardin N, Briani G, Mostefaoui S, Morinaud G, Crane B, Szwec N, Delauche L, Jamme F, Sandt C, Dumas P (2013) Ultracarbonaceous Antarctic micrometeorites, probing the solar system beyond the nitrogen snow-line. *Icarus* 224:243–252
- Duprat J, Dobrica E, Engrand C, Aléon J, Marrocchi Y, Mostefaoui S, Meibom A, Leroux H, Rouzaud J-N, Gounelle M, Robert F (2010) Extreme deuterium excesses in ultracarbonaceous micrometeorites from central Antarctic snow. *Science* 328:742–745

- Ehrenfreund P, Charnley SB (2000) Organic molecules in the interstellar medium, comets, and meteorites: a voyage from dark clouds to the early Earth. *Annu Rev Astron Astrophys* 38:427–483
- Floss C, Noguchi T, Yada T (2012) Ultracarbonaceous Antarctic micrometeorites: origins and relationships to other primitive extraterrestrial materials. *Lunar Planet Sci XLIII*. #1217
- Fujiya W, Sugiura N, Sano Y, Hiyagon H (2013) Mn–Cr ages of dolomites in CI chondrites and the Tagish Lake ungrouped carbonaceous chondrite. *Earth Planet Sci Lett* 362:130–142
- Glavin DP, Dworkin JP, Sandford SA (2008) Detection of cometary amines in samples returned by stardust. *Meteorit Planet Sci* 43:399–413
- Glavin DP, Alexander CMO'D, Aponte JC, Dworkin JP, Elsila JE, Yabuta H (2018) The origin and evolution of organic matter in carbonaceous chondrites and links to their parent bodies. In: Abreu N (ed) *Primitive meteorites and asteroids: physical, chemical, and spectroscopic observations paving the way to exploration*. Elsevier, Amsterdam, pp 205–271
- Greenberg JM, Li A (1997) Silicate core-organic refractory mantle particles as interstellar dust and as aggregated in comets and stellar disks. *Adv Space Res* 19:981–990
- Hill HGM, Nuth JA (2003) The catalytic potential of cosmic dust: implications for prebiotic chemistry in the solar nebula and other protoplanetary systems. *Astrobiology* 3:291–304
- Huss GR, Rubin AE, Grossman JN (2006) Thermal metamorphism in chondrites. In: Lauretta DS, McSween HY Jr (eds) *Meteorites and the early solar system II*. University of Arizona Press, Tucson, pp 567–586
- Kebukawa Y, Kilcoyne ALD, Cody GD (2013) Exploring the potential formation of organic solids in chondrites and comets through polymerization of interstellar formaldehyde. *Astrophys J* 771:19
- Keller LP, Messenger S (2011) On the origins of GEMS grains. *Geochim Cosmochim Acta* 75:5336–5365
- Kuga M, Marty B, Marrocchi Y, Tissandier L (2015) Synthesis of refractory organic matter in the ionized gas phase of the solar nebula. *Proc Natl Acad Sci U S A* 112:7129–7134
- Lauretta DS, Balram-Knutson SS, Boynton BEWV, Drouet d'Aubigny C, Dellagiustina DN, Enos HL, Golith DR, Hergenrother CW, Howell ES, Bennett CA, Morton ET, Nolan MC, Rizk B, Roper HL, Bartels AE, Bos BJ, Dworkin JP, Highsmith DE, Lorenz DA, Lim LF, Mink R, Moreau MC, Nuth JA, Reuter DC, Simon AA, Bierhaus EB, Bryan BH, Ballouz R, Barnouin OS, Binzel RP, Bottke WF, Hamilton VE, Walsh KJ, Chesley SR, Christensen PR, Clark BE, Connolly HC, Crombie MK, Daly MG, Emery JP, McCoy TJ, McMahon JW, Scheeres DJ, Messenger S, Nakamura-Messenger K, Righter K, Sandford SA (2017) OSIRIS-REx sample return from asteroid (101955) Bennu. *Space Sci Rev*. <https://doi.org/10.1007/s11214-017-0405-1>
- McKeegan KD, Aléon J, Bradley J, Brownlee D, Busemann H, Butterworth A, Chaussidon M, Fallon S, Floss C, Gilmour J, Gounelle M, Graham G, Guan Y, Heck PR, Hoppe P, Hutcheon ID, Huth J, Ishii H, Ito M, Jacobsen SB, Kearsley A, Leshin LA, Liu M-C, Lyon I, Marhas K, Marty B, Matrajt G, Meibom A, Messenger S, Mostefaoui S, Mukhopadhyay S, Nakamura-Messenger K, Nittler L, Palma R, Pepin RO, Papanastassiou DA, Robert F, Schlutter D, Snead CJ, Stadermann FJ, Stroud R, Tsou P, Westphal A, Young ED, Ziegler K, Zimmermann L, Zinner E (2006) Isotopic compositions of cometary matter returned by stardust. *Science* 314:1724–1728
- Meier R, Owen TC, Jewitt DC, Matthews HM, Senay M, Biver N, Bockelée-Morvan D, Crovisier J, Gautier D (1998) Deuterium in Comet C/1995 O1 (Hale-Bopp). Detection of DCN. *Science* 279:1707–1710
- Messenger S (2000) Identification of molecular-cloud material in interplanetary dust particles. *Nature* 404:968–971
- Millar TJ, Bennett A, Herbst E (1989) Deuterium fractionation in dense interstellar clouds. *Astrophys J* 340:906–920
- Nakamura T, Noguchi T, Ozono Y, Osawa T, Nagao K (2005) Mineralogy of ultracarbonaceous large micrometeorites. *Meteor Planet Sci* 40:A110

- Nakamura-Messenger K, Messenger S, Keller LP, Clemett SJ, Zolensky ME (2006) Organic globules in the Tagish Lake meteorite: remnants of the protosolar disk. *Science* 314:1439–1442
- Nakamura-Messenger K, Clemett SJ, Messenger S, Keller LP (2011) Experimental aqueous alteration of cometary dust. *Meteor Planet Sci* 46:843–856
- Noguchi T, Ohashi N, Tsujimoto S, Mitsunari T, Bradley JP, Nakamura T, Toh S, Stephan T, Iwata N, Imae N (2015) Cometary dust in Antarctic ice and snow: past and present chondritic porous micrometeorites preserved on the Earth's surface. *Earth Planet Sci Lett* 410:1–11
- Noguchi T, Yabuta H, Itoh S, Sakamoto N, Mitsunari T, Okubo A, Okazaki R, Nakamura T, Tachibana S, Terada K, Ebihara M, Imae N, Kimura M, Nagahara H (2017) Variation of mineralogy and organic material during the early stages of aqueous activity recorded in Antarctic micrometeorites. *Geochim Cosmochim Acta* 208:119–144
- Nomura H, Aikawa Y, Tsujimoto M, Nakagawa Y, Millar TJ (2007) Molecular hydrogen emission from protoplanetary disks: effects of X-ray irradiation and dust evolution. *Astrophys J* 661:334–353
- Nuevo M, Milam SN, Sandford SA, De Gregorio BT, Cody GD, Kilcoyne ALD (2011) XANES analysis of organic residues produced from the UV irradiation of astrophysical ice analogs. *Adv Space Res* 48:1126–1135
- Öberg KI (2016) Photochemistry and astrochemistry: photochemical pathways to interstellar complex organic molecules. *Chem Rev* 116:9631–9663
- Pizzarello S, Cooper GW, Flynn GJ (2006) The nature and distribution of the organic material in carbonaceous chondrites and interplanetary dust particles. In: Luretta DS, McSween HY (eds) *Meteorites and the early solar system II*. The University of Arizona Press, Tucson, pp 625–651
- Pizzarello S, Williams LB, Lehman J, Holland GP, Yarger JL (2011) Abundant ammonia in primitive asteroids and the case for a possible exobiology. *Proc Natl Acad Sci U S A* 108:4303–4306
- Quirico E, Moroz LV, Schmitt B, Arnold G, Faure M, Beck P, Bonal L, Ciarniello M, Capaccioni F, Filacchione G, Erard S, Leyrat C, Bockelée-Morvan D, Zinzi A, Palomba E, Drossart P, Tosi F, Capria MT, De Sanctis MC, Raponi A, Fonti S, Mancarella F, Orofino V, Barucci A, Blecka MI, Carlson R, Despan D, Faure A, Fornasier S, Gudipati MS, Longobardo A, Markus K, Mennella V, Merlin F, Piccioni G, Rousseau B, Taylor F, Rosetta VIRTIS team (2016) Refractory and semi-volatile organics at the surface of comet 67P/Churyumov Gerasimenko: insights from the VIRTIS/Rosetta imaging spectrometer. *Icarus* 272:32–47
- Sandford SA (1996) The inventory of interstellar materials available for the formation of the solar system. *Meteorit Planet Sci* 31:449–476
- Sandford SA, Aléon J, Alexander CMO'D, Araki T, Bajt S, Baratta GA, Borg J, Bradley JP, Brownlee DP, Brucato JR, Burchell MJ, Busemann H, Butterworth A, Clemett SJ, Cody G, Colangeli L, Cooper G, D'Hendecourt L, Djouadi Z, Dworkin JP, Ferrini G, Fleckenstein H, Flynn GJ, Franchi IA, Fries M, Gilles MK, Glavin DP, Gounelle M, Grossemy F, Jacobsen C, Keller LP, Kilcoyne ALD, Leitner J, Matrajt G, Meibom A, Mennella V, Mostefaoui S, Nittler LR, Palumbo ME, Papanastassiou DA, Robert F, Rotundi A, Snead CJ, Spencer MK, Stadermann FJ, Steele A, Stephan T, Tsou P, Tylliszczak T, Westphal AJ, Wirick S, Wopenka B, Yabuta H, Zare RN, Zolensky M (2006) Organics captured from comet 81P/wild 2 by the stardust spacecraft. *Science* 314:1720–1724
- Schmitt-Kopplin P, Gabelica Z, Gougeon RD, Fekete A, Kanawati B, Harir M, Gebeuegi I, Eckel G, Hertkorn N (2010) High molecular diversity of extraterrestrial organic matter in Murchison meteorite revealed 40 years after its fall. *Proc Natl Acad Sci U S A* 107:2763–2768
- Schulz R, Jehin E, Manfroid J, Hutsemékers D, Arpigny C, Cochran A, Zucconi J, Stuwe J (2008) Isotopic abundance in the CN coma of comets: ten years of measurements. *Planet Space Sci* 56:1713–1718
- Sqyres SW, Nakamura-Messenger K, Mitchell DF, Moran VE, Houghton MB, Glavin DP, Hayes AG, Lauretta DS, the CAESAR Project Team (2018) The CAESAR new frontiers mission: 1. Overview. The 49th Lunar and Planetary Science conference, Abstract#1332
- Stöhr J (1991) NEXAFS spectroscopy. Springer, New York. 291 p

- Tachibana S, Abe M, Arakawa M, Fujimoto M, Iijima Y, Ishiguro M, Kitazato K, Kobayashi N, Namiki N, Okada T, Okazaki R, Sawada H, Suguta S, Takano Y, Tanaka S, Watanabe S, Yoshikawa M, Kuninaka H, the Hayabusa2 Project Team (2014) Hayabusa2: scientific importance of samples returned from C-type near-Earth asteroid (162173) 1999 JU(3). *Geochem J* 48:571–587
- Tholen DJ, Barucci MA (1989) Asteroid taxonomy. In: Binzel RP, Gehrels T, Matthews MS (eds) *Asteroids II*. University of Arizona Press, Tucson, pp 298–315
- Tielen AGHM, Hagen W (1982) Model calculations of the molecular composition of interstellar grain mantles. *Astron Astrophys* 114:245–260
- Walsh KJ, Morbidelli A, Raymond SN, O’Brien DP, Mandell AM (2012) Populating the asteroid belt from two parent source regions due to the migration of giant planets—“The Grand Tack”. *Meteorit Planet Sci* 47:1–7
- Yabuta H (2018) Solar system exploration: small bodies and their chemical and physical conditions. In: Kolb V (ed) *Handbook of astrobiology*. CRC Press, Boca Raton
- Yabuta H, Noguchi T, Itoh S, Nakamura T, Miyake A, Tsujimoto S, Ohashi N, Sakamoto N, Hashiguchi M, Abe K-I, Okubo A, Kilcoyne ALD, Tachibana S, Okazaki R, Terada K, Ebihara M, Nagahara H (2017) Formation of an ultracarbonaceous Antarctic micrometeorite through minimal aqueous alteration in a small porous icy body. *Geochim Cosmochim Acta* 214:172–190
- Yabuta H, Sandford SA, Meech KJ (2018) Organic molecules and volatiles in comets. *Elements* 14:101–106
- Yada T, Floss C, Stadermann FJ, Zinner E, Nakamura T, Noguchi T, Lea S (2008) Stardust in Antarctic micrometeorites. *Meteorit Planet Sci* 43:1287–1298
- Yano H, Sasaki S, Imani J, Horikawa D, Arai K, Fujishima K, Hashimoto H, Higashide M, Imai E, Ishibashi Y, Kawaguchi Y, Kawai H, Kebukawa Y, Kobayashi K, Kobunai K, Kodaira S, Kurosu Y, Mita H, Oda Y, Okudaira K, Ozawa T, Tabata M, Takizawa N, Tomita M, Tsuchiyama A, Uchihori Y, Yabuta H, Yaguchi Y, Yokobori S, Yamagishi A, the Tanpopo Project Team (2017) In-orbit operation and initial sample analysis and curation results for the first year collection samples of the Tanpopo project. The 48th Lunar and Planetary Science conference, Abstract#3040

Chapter 4

Prebiotic Synthesis of Bioorganic Compounds by Simulation Experiments



Kensei Kobayashi

Abstract A great number of prebiotic synthesis experiments under possible primitive conditions have been conducted since the 1950s. In most of the prebiotic synthesis experiments of the earlier era, strongly reducing gas mixtures were used as a primitive Earth atmosphere, and amino acids and other bioorganic compounds were successfully synthesized. Their formation mechanisms were explained by step-by-step modes, such as the Strecker synthesis. However, such a strongly reducing atmosphere is questioned and the contributions of extraterrestrial organics are under consideration. We learned that quite complex organic compounds could be formed under interstellar environments through the analysis of extraterrestrial samples and products of experiments simulating extraterrestrial conditions. Simulation experiments are also introduced to examine the possible origins of the homochirality of biomolecules. Experiments simulating submarine hydrothermal systems were also conducted. It is very difficult to verify the origin of life on the Earth, since relics of the prebiotic synthesis do not remain on the Earth. It would be possible, however, to examine possible origins of life in space through the synergy of planetary exploration and space experiments.

Keywords Prebiotic synthesis · Amino acids · Extraterrestrial organics · Homochirality · Submarine hydrothermal systems

4.1 Introduction

Chemical evolution is the concept used to explain that the first life was abiotically generated by the evolution of simple molecules into complicated and systemized organic compounds. Oparin and Haldane first presented this idea independently in the 1920s (Oparin 1953; Haldane 1929). It was assumed that it would be difficult to verify the chemical evolution processes experimentally in their era. In the early 1950s, however, pioneering experiments to examine chemical evolution pathways,

K. Kobayashi (✉)
Yokohama National University, Yokohama, Japan
e-mail: kobayashi-kensei-wv@ynu.ac.jp

including the historical spark discharge experiment by Miller (1953), were carried out. Following this experiment, a great number of experimental studies have been conducted to examine possible prebiotic chemical evolution pathways. Such works included experiments simulating the primitive Earth atmosphere, the primeval ocean, interstellar environments, and so on.

In this chapter, such experiments that were done under simulated primitive environments are introduced with a focus on our experiments, and a future perspective will be also provided.

4.2 Dawn of Experimental Prebiotic Chemistry

4.2.1 *One-Pot Reactions*

In the 1920s, Oparin and Haldane presented the theoretical idea of chemical evolution, but it was assumed that experimental verification of the idea would be quite difficult since such an “evolution” would require far greater time than could be provided in the laboratories of their period.

In the middle of the twentieth century, there were two hypotheses on the composition of the primitive Earth atmosphere, namely, a strongly reducing one and a nonreducing one. The former states that methane and ammonia were major constituents of the atmosphere, which was modeled based on the cosmic abundance and the atmospheres of the giant planets in our solar system (Urey 1952). In the latter theory, carbon dioxide and nitrogen were considered to be major constituents of the atmosphere, as seen on present terrestrial planets like Venus and Mars (Abelson 1966).

Garrison et al. (1951) performed simulation experiments based on the latter (nonreducing) atmospheric model. They irradiated aqueous solutions containing CO_2 and Fe^{2+} with high-energy helium ions from an accelerator to simulate the action of alpha rays resulting from radioactive nuclides in the primeval ocean. They detected formaldehyde (HCHO) and formic acid (HCOOH) in the products, but could not find evidence of important bioorganic chemicals such as amino acids since the system did not contain nitrogen.

Two years later, Miller (1953) followed the ideas of Urey (his supervisor) in that the primitive Earth atmosphere was strongly reducing and performed historical spark discharge experiments in a gas mixture of methane, ammonia, hydrogen, and water. In his first paper, five amino acids, namely, glycine, alanine, aspartic acid, β -alanine, and α -aminobutyric acid, were detected in the aqueous products by paper chromatography.

Since it was so astonishing that amino acids, most important bioorganics, were formed easily from simple molecules, many scientists started research in the field of experimental prebiotic chemistry. Most of these scientists used the strongly reduc-

ing gas mixtures containing methane and ammonia, which were exposed to various energy sources (Miller and Orgel 1974). They simulated the roles of solar ultraviolet light (Sagan and Khare 1971), thermal energy from volcanoes (Harada and Fox 1964), and shock waves associated with meteor impacts (Bar-Nun et al. 1970). All of these experiments resulted in the detection of amino acids in the products of these reactions. It was shown that formation of amino acids was not difficult to achieve if the primitive Earth had a strongly reducing atmosphere. On the other hand, nonreducing gas mixtures without methane and ammonia did not produce amino acids (Schlesinger and Miller 1983). Many people in the field therefore preferred the theory of a strongly reducing atmosphere to a nonreducing one. The composition of Earth's primitive atmosphere will be discussed in Chap. 10.

Nucleic acid bases are another group of important bioorganics. Oro (1960) successfully synthesized adenine in an aqueous solution of hydrogen cyanide and ammonia. Ponnampetuma et al. (1963) found adenine when they performed an electron irradiation of a mixture of methane, ammonia, and water. It was suggested that adenine, among the nucleic acid bases, was abiotically formed most easily in reducing environments since adenine ($C_5H_5N_5$) is a more reducing base than guanine ($C_5H_5ON_5$), cytosine ($C_4H_5ON_3$), uracil ($C_4H_4O_2N_2$), and thymine ($C_5H_7O_2N_2$).

Sugars (carbohydrates) could be formed from a formaldehyde aqueous solution with a mild alkaline catalyst via a reaction known as the formose reaction (Butlerow 1861). There are, however, problems in synthesizing ribose, the sugar used in RNA, using the formose reaction under prebiotic conditions. These issues stem from the required high concentration of formaldehyde, and the yield of ribose was generally low, since so many other kinds of sugars and sugar-like compounds were formed in it.

4.2.2 Energetics and Formation Mechanisms

Table 4.1 summarizes the various energy sources available on the primitive Earth (Kobayashi and Saito 2000). Among them, the solar ultraviolet light flux is notably far larger than the others, followed by lightning (electric discharges). Amino acids could be formed by any of the energy sources listed here, and it was thus suggested that predominant energies such as solar UV and lightning were important in the abiotic synthesis of bioorganic compounds on the primitive Earth.

Miller and Urey suggested that amino acids were formed via the Strecker-type reactions in a spark discharge flask since they found hydrogen cyanide and aldehydes in their discharge products. The Strecker synthesis is a famous organic synthesis technique used to form amino acids, where hydrogen cyanide (HCN), aldehydes (RCHO), and ammonia (NH_3) are used as starting materials. These are mixed to form aminonitriles ($NH_2-CHR-CN$), and the subsequent hydrolysis of these aminonitriles produces amino acids ($NH_2-CHR-COOH$).

Table 4.1 Major energy sources for prebiotic synthesis on the primitive Earth

Energy source	Estimated flux /eV m ⁻² year ⁻¹	Reference	Amino acid formation	
			Strongly-reducing ^a	Mildly-reducing ^b
Solar radiation				
Total	6.8 × 10 ²⁸	1		
λ < 200 nm	2.2 × 10 ²⁵	1	+++	–
λ < 150 nm	9.1 × 10 ²³	1	+	–
λ < 110 nm	4.2 × 10 ²²	2	+	+
Thundering	1.8 × 10 ²²	3	++	+
	1.0 × 10 ²⁴	1		
Volcano heat	3.4 × 10 ²²	1	+	–
Radioactivity ^c	2 × 10 ²³	1	+	+
Cosmic rays	2.9 × 10 ²¹	2	+	+++
Meteor impacts	1 × 10 ²²	1	+	++

Reference 1. Miller and Urey (1959), 2. Kobayashi et al. (1998), 3. Chyba and Sagan (1991)

^aCH₄-NH₃-H₂O atmosphere

^bCO-N₂-H₂O atmosphere

^c0–1.0 km deep of Earth crust

4.2.3 Step-by-Step Reaction Models

Simulation experiments using gas mixtures and various energy sources can be referred to one-pot reactions. Oro's adenine synthesis could be categorized as a one-pot reaction. Other types of experiments were also conducted on this subject. These used a selected number of starting compounds that were assumed to be present in prebiotic conditions as suggested by previous studies.

Not only was adenine detected upon the use of these experiments, but guanine was also detected by HCN polymerization (Levy et al. 1999). Pyrimidine bases (including uracil, cytosine, and thymine), however, could not be obtained from the HCN solution. Ferris et al. (1968) detected cytosine following a reaction between cyanoacetylene (CHCCN) with cyanate (HCNO). Uracil was obtained by the hydrolysis of cytosine (Miller and Orgel 1974). Thus it was noted that cyanoacetylene, not HCN, could be a candidate starting material for the production of pyrimidine bases.

After obtaining purines, pyrimidines, and sugars (ribose), the next targets for synthesis were nucleosides and nucleotides. Several studies on such syntheses were performed in the 1960s–1970s. Fuller et al. (1972) heated and dried a mixture containing a purine base (adenine, guanine, or hypoxanthine), ribose, magnesium ions, and trimetaphosphate and obtained purine nucleosides. Lohrmann and Orgel (1971) heated a mixture of uridine, inorganic phosphates, urea, and ammonium ions and obtained uridine phosphate. In both cases, these products were mixtures of isomers (e.g., α-nucleoside and β-nucleoside).

The oligomerization of amino acids and nucleosides has also been studied. Oligomerization is a condensation reaction and requires the removal of water

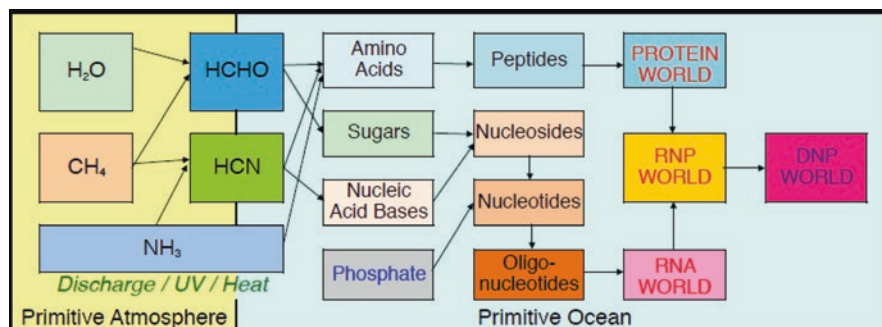


Fig. 4.1 Conventional stepwise scenario of chemical evolution

molecules from monomers. When amino acids were heated to dryness, peptides were often detected (e.g., Harada and Fox 1960). In the case of oligonucleotides, mononucleotides should have been activated. Phosphoimidazolide-activated nucleotides (ImpN) were often used in conjunction with metal ions and/or templates, and oligonucleotides were thus obtained (e.g., Sawai 1976).

After a great number of laboratory simulation experiments, a scenario outlining the processes from simple molecules to the existence of life was roughly expressed, as shown in Fig. 4.1. Here it may be referred to as a conventional stepwise scenario. Please note that each step was confirmed by the use of some *in vitro* simulation experiments, where high concentrations of starting materials were used without adding any inhibitors. Recent progresses in prebiotic syntheses of RNA will be introduced in Chap. 5.

4.3 Formation and Delivery of Extraterrestrial Organic Compounds

In the 1970s, it was recognized that ammonia in a terrestrial atmosphere was easily decomposed by solar UV radiation. A mixture of methane and nitrogen (N_2) was often used in experiments simulating abiotic synthesis in the primitive Earth atmosphere, and amino acids were still formed using such energies as spark discharges (Ring et al. 1972; Kobayashi and Ponnampertuma 1985b). However, the formation of amino acids by solar UV radiation was not achieved, since nitrogen cannot be dissociated by near UV light (Ferris and Chen 1975; Bar-Nun and Chang 1968). In the 1980s, planetary explorations of the solar system provided new insight into the formation of planets: they were formed by the collision of planetesimals, which lead to the formation of mildly reducing secondary planetary atmospheres (Abe and Matsui 1986a, b, Kasting 1990; Catling and Kasting 2017). Schlesinger and Miller (1983) performed spark discharge experiments using $CH_4-N_2-H_2O$, $CO-N_2-H_2O$, and $CO_2-N_2-H_2O$ type gas mixtures, and found that mixtures containing CO or CO_2

produced a much smaller amount of amino acids than CH_4 . Thus it is supposed that the formation of amino acids and other bioorganic compounds in the mildly reducing atmosphere is limited.

About the same period, a wide variety of organic compounds were found in extraterrestrial environments, such as in meteorites and comets. Kvenvolden et al. (1970) found amino acids in the Murchison meteorite, which was a CM2 carbonaceous chondrite that fell in Australia in 1969. The amino acids found in it were racemic mixtures, demonstrated their indigenousness. Complex organic compounds were also detected in cometary dusts by impact ionization mass spectrometry during the Vega 1 mission to Comet Halley (Kissel and Krueger 1987). Glycine was detected in a sample from Comet Wild 2 that was returned by the Stardust spacecraft (Elsila et al. 2009). These findings supported the formation of organic compounds in space and transfer to the Earth.

4.3.1 Abiotic Syntheses of Amino Acids in Simulated Space Environments

There are a number of scenarios that may be used to explain how organic compounds found in meteorites and comets formed, including formation in interstellar grains and in meteorite parent bodies. The former was first proposed by Greenberg and Li (1997) as follows (Fig. 4.2): interstellar molecules including H_2O , CO , CH_3OH , and NH_3 are frozen into silicate dusts that form ice mantles in molecular clouds since the temperature inside of the molecular clouds is quite low. The ice mantles are then irradiated by cosmic rays and cosmic ray-induced ultraviolet light, and complex organic molecules are thus formed. Such organic materials were

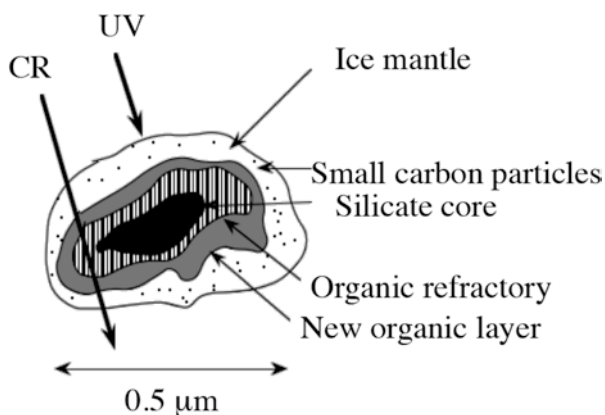


Fig. 4.2 Formation of organic compounds in interstellar dust environments. (Modified from Greenberg and Mendoza-Gomez 1993)

brought into asteroids and comets when they were formed from dusts in the early solar system.

Though it is possible to observe simple molecules such as CO and CH₃OH in the ice mantles of molecular clouds, it is not easy to detect complex molecules. Instead, a number of laboratory simulation experiments have been conducted to see what kinds of organic molecules are formed in the ice mantles. The first report of amino acid formation in simulated ice mantles was reported by Briggs et al. (1992). They irradiated a frozen mixture of H₂O, CO, and NH₃ on a metal substrate using UV light from a hydrogen lamp at 12 K. After irradiation, they warmed up the substrate to room temperature and analyzed the product residue using GC/MS. They detected glycine as well as a number of other organic molecules.

Kobayashi et al. examined the possible roles of cosmic ray particles in place of ultraviolet photons in the formation of complex organic molecules in the ice mantles. Ice mixtures of H₂O, CO, and NH₃ on metal substrates were irradiated with 3 MeV protons from a van de Graaff accelerator (Tokyo Institute of Technology) (Fig. 4.3), and it was found that several amino acids existed in the hydrolyzed products (Kobayashi et al. 1995; Kasamatsu et al. 1997). The energy of the cosmic rays is considered too high and passes through the ice mantles with little interaction with other molecules. Kobayashi et al. (2010) used 290 MeV/u carbon beams, whose energy is close to typical cosmic rays that exist in space, which were generated by HIMAC accelerator (National Institute of Radiological Sciences). A frozen mixture of H₂O, CH₃OH, and NH₃ at 77 K was irradiated with carbon beams, and the resulting product was acid hydrolyzed and was subjected to amino acid analysis. A number of amino acids were detected by HPLC and/or GC/MS. It was shown that cosmic ray particles, which deposit only a small part of their energy, can form complex organic molecules including amino acid precursors.

Since UV sources are more easily available than accelerators, more UV irradiation experiments have been performed than particle irradiation experiments, and

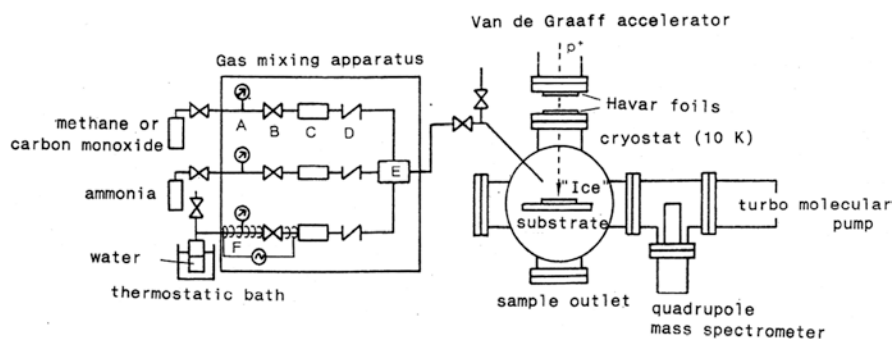


Fig. 4.3 Proton irradiation of a simulated ice mantle of interstellar dust particles (ISDs). (a) Pressure gauge, (b) needle valve, (c) thermal mass-flow meter, (d) control valve, (e) mixer. A gas mixture of CO (or CH₄), NH₃, and H₂O was frozen on a metal substrate located in a cryostat and was irradiated with high-energy protons from a van de Graaff accelerator via Havar foils. (Modified from Kobayashi et al. 1995)

many kinds of racemic amino acids were identified in the irradiation products after hydrolysis (Munos Caro et al. 2002; Bernstein et al. 2002). These experiments support the formation of amino acids in space environment.

4.3.2 Abiotic Syntheses of Amino Acids and Insoluble Organic Matter in Simulated Meteorite Parent-Body Environments

Most of the organic carbon found in carbonaceous chondrites is often referred to as IOM (insoluble organic matter), which has quite complex macromolecular structure (Pizzarello 2006). One of the possible scenarios of the formation of IOM is via aqueous alteration or a hydrothermal reaction in the parent bodies of the meteorites. Cody et al. (2011) proposed aqueous solution of formaldehyde could yield insoluble organic matter since water ice melted via heat provided by the decay of ^{26}Al . The IOM obtained by their experiments had a structure similar to the IOM found in carbonaceous chondrites. Further studies showed that the incorporation of ammonia in this reaction could produce organic solids containing imidazole, pyridine, and pyrrole structures (Kebukawa et al. 2013).

Amino acids were also found from similar experiments. When a mixture of formaldehyde, glycolaldehyde (CH_2OHCHO), ammonia, and water with an alkaline catalyst ($\text{Ca}(\text{OH})_2$) was heated at 90–250 °C for 72 h, both water-soluble and insoluble organic compounds were formed. The soluble fraction produced amino acids after acid hydrolysis (Kebukawa et al. 2017). It was shown that both IOM and amino acid precursors could be formed in the interior of meteorite parent bodies in the early stages of the solar system.

4.3.3 Formation of Enantiomeric Excesses of Amino Acids in Extraterrestrial Environments

Terrestrial organisms generally use L-amino acids (except achiral glycine) to synthesize proteins, while abiotically formed amino acids were fundamentally racemic mixtures, i.e., a 1:1 mixture of D- and L-amino acids (Fig. 4.4). Why then were only L-amino acids selected when life was born on the Earth? The riddle of the origin of the homochirality of these amino acids (and other biomolecules) has been one of the largest subjects in the study of chemical evolution toward the generation of life. Though a number of hypotheses for this have been presented (Bonner 1991), most of them have not been supported by cosmochemical evidence.

Amino acids found in meteorites were thought to be fundamentally racemic mixtures since they are formed abiotically. Cronin and Pizzarello (1997) reported that some of the amino acids found in the Murchison meteorite had L-enantiomer excesses. Though L-excesses of protein amino acids have been judged to stem from

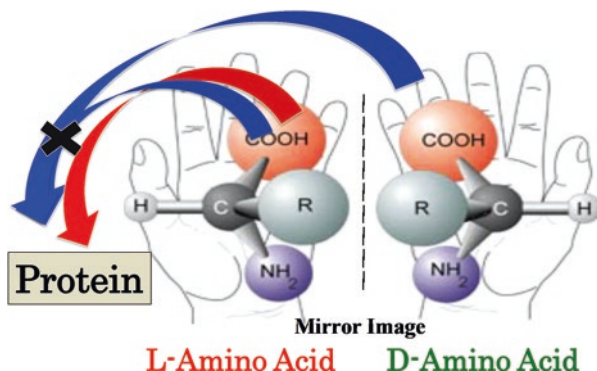


Fig. 4.4 Amino acid enantiomers. (Modified from Kobayashi et al. 2010)

contamination from the terrestrial biosphere, this group showed that α -methyl amino acids such as isovaline also have some L-excesses. This tendency has been confirmed by several investigators (Elsila et al. 2016). It was suggested that these enantiomeric excesses (ee's) of amino acids were generated by physical asymmetry in space. Among several candidates, circularly polarized ultraviolet light (CPL-UV) has been considered most promising as the cause of this physically asymmetric formation. CPL-UV from neutron stars was first noted as a potential cause, but there is little chance to be exposed to CPL-UV, since it is highly directional. On the other hand, widely distributed CPL regions were recently found (Fukue et al. 2010), which are more plausible triggers for asymmetric formation of amino acids in space.

The generation of ee's of amino acids by CPL-UV has been often explained by asymmetric decomposition. One of the amino acid enantiomers decomposes more than the other after irradiation of the left- or right-handed CPL-UV, which was shown theoretically and experimentally (Inoue 1992). A problem in this scenario is that UV photons may have decomposed most of the amino acids to obtain significant ee's of the amino acids resulting in a few remaining quantity.

Takano et al. (2007) irradiated not free amino acids, but amino acid precursors using CPL from a synchrotron. Amino acid precursors were synthesized from CO, NH₃, and H₂O by proton irradiation, which were quite complex organic molecules (Takano et al. 2004). After right-CPL was irradiated on the complex amino acid precursors, a small amount (0.44%) of ee's of D-alanine was detected, while the left-CPL resulted in 0.65% of ee's of L-alanine. The detected ee's were small but statistically significant. It is notable that the yield of amino acids after hydrolysis following CPL-UV irradiation was not greatly changed from unirradiated precursors. It can be shown that the amino acids were not asymmetrically decomposed but rather asymmetrically altered (Fig. 4.5).

Marcelus et al. (2011) also found such ee's after a frozen mixture of CH₃OH, NH₃, and H₂O (80 K) was irradiated with CPL-UV. It was found that the amino acid

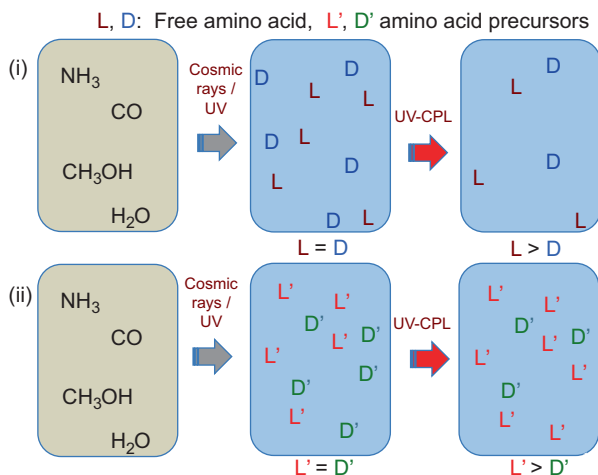


Fig. 4.5 Generation of enantiomeric excesses of amino acids in space. (i) Conventional asymmetric decomposition of free amino acids. More free D-amino acids are decomposed by irradiation with circularly-polarized light (CPL) than free L-amino acids. (ii) Newly proposed asymmetric alteration of amino acid precursors. Complex organics were formed in interstellar space having both L- and D-amino acid moieties equally. A part of the D-amino acid moieties were converted to the L-amino acid moieties by irradiation with CPL

precursors were formed in the ice by UV irradiation, and further CPL-UV radiation at room temperature formed 0.71–1.34% ee's.

In the CPL-UV-triggered chirogenesis hypothesis, the chance to generate an L-amino acid favoring world is 50%, since there are both left-CPL regions and right-CPL regions in space (Fukue et al. 2010). Another group of hypotheses are those based on the parity violation. One example of the parity violation is that electrons generated during β -decays are always left-spin polarized. It is expected that spin-polarized electrons can induce a similar effect as CPL on amino acid decomposition/alteration. Compared to the CPL experiments, spin-polarized electron experiments are difficult. One possible way to examine the effects of these electrons is to use natural β -ray sources, which produce left spin-polarized electrons. When racemic mixtures of amino acids were irradiated with β -rays from a strong ^{90}Sr - ^{90}Y source, optical activity was observed in the irradiated samples, but not in the samples taken before irradiation (Burkov et al. 2008; Gusev et al. 2008). Further experiments are needed in this field, particularly those with right spin-polarized electrons.

4.3.4 Delivery of Extraterrestrial Organic Compounds

We are now sure that there are a wide variety and a large amount of organic compounds including amino acid precursors and nucleic acid bases in space. The next question is how to deliver these compounds to the primitive Earth. Though we know that some carbonaceous chondrites have delivered organic compounds safely to the Earth, organics in larger meteorites would have decomposed during bolide impacts. It was suggested that cosmic dusts could have delivered more organics than meteorites and comets (Chyba and Sagan 1992; Barbier et al. 1998). Cosmic dusts are, however, so tiny that they are exposed to strong solar UV radiation during their stay in space. Cosmic dusts (micrometeorites) have been collected in the terrestrial biosphere such as in ice from Antarctica, and they have been found to contain a high percentage of organics, but most of these are complex insoluble organic matter (IOM). It is difficult to conclude that they could carry such bioorganics as amino acids since they have been in contact with terrestrial materials.

Yamagishi et al. (2009) have been conducting a space experiment named Tanpopo by utilizing the International Space Station since 2015. In this mission, super low-density silica aerogel is exposed on the Exposed Facility of the Japanese Experiment Module (JEM-EF) and is collecting dusts flying near the ISS. Amino acids from the captured dusts will be analyzed to examine the scenario that extraterrestrial amino acids and other bioorganics were delivered by cosmic dusts.

4.4 Abiotic Synthesis and Alteration of Organic Compounds in Simulated Primitive Earth Environments

Though organic compounds including amino acid precursors could have been formed either in interstellar space or in planetary atmospheres, life cannot be generated there. Since water is essential for life and a major molecule consisting terrestrial organisms is water, it is natural that terrestrial life was first generated in a hydrosphere – either in the ocean or in land water. Cold or warm (up to 100 °C) water media having plenty of sunlight were considered as sites of the generation of life on the Earth up until the 1970s.

4.4.1 Prebiotic Synthesis in Simulated Submarine Hydrothermal Conditions

Corliss et al. (1979) first discovered submarine hydrothermal vents at the Galapagos spreading center. Superheated seawater over 300 °C was erupting from chimneys on the deep seafloor. Such submarine hydrothermal systems have a number of merits for their use as sites for chemical evolution toward the origin of life: (1) the

hydrothermal fluid contains such reducing species as hydrogen, methane, and ammonia, maintaining reducing environments even at present on Earth; (2) seawater is superheated by magma and then quenched when it erupts into cold seawater; (3) the hydrothermal fluid contains high concentrations of essential metal ions such as iron, manganese, and zinc, which can work as biotic and prebiotic catalysts (Kobayashi and Ponnampertuma 1985a, b); (4) the deep sea is a dark world, where organic compounds were free from decomposition from strong solar UV radiation before the formation of the ozone layer. It was also suggested that the last universal common ancestor (LUCA) or commonote of terrestrial organisms were hyperthermophiles, as suggested by the 16S rRNA universal phylogenetic tree (Pace 1991), which agrees with the idea that the first terrestrial life was generated near superheated media (see also Chap. 7).

In the earlier stages of experiments simulating submarine hydrothermal systems, autoclaves were often used, where possible starting materials were dissolved in water and heated at a high pressure following pressurization. For example, Yanagawa and Kobayashi (1992) heated an aqueous solution containing various metal ions (Fe^{2+} , Mn^{2+} , Zn^{2+} , Ca^{2+} , Cu^{2+} , and Ba^{2+}) and NH_4^+ at 325 °C for 1.5–12 h in an autoclave under pressurization by an 80 kg cm⁻² gas mixture of CH_4 and N_2 . The resulting products contained amino acids following hydrolysis.

Imai et al. (1999) developed a flow reactor simulating submarine hydrothermal systems. An aqueous solution of glycine was injected into the reactor and heated at 250 °C and then quenched at 0 °C, which yielded a hexamer of glycine in the solution. Kurihara et al. (2012) examined the possible formation of organic aggregates in simulated submarine hydrothermal systems. First a mixture of CO , N_2 , and H_2O was irradiated with high-energy protons to simulate possible organic formation in a primitive atmosphere. Water-soluble complex organic compounds containing amino acid precursors were formed. The products were heated at 300 °C under a pressure of 25 MPa and then quenched to 0 °C in a flow reactor. The resulting products contained organic aggregates with amino acid precursors. These results suggested that it was possible that in submarine hydrothermal systems, the dehydration condensation of biomonomers and the formation of organic aggregates in an aqueous solution had taken place.

4.4.2 *Reconsideration of the Stepwise Scenario of Chemical Evolution*

It has been controversial as to whether proteins and RNA were formed earlier than other organic materials. After the discovery of ribozymes (Krueger et al. 1982), Gilbert (1984) proposed the hypothesis that the first terrestrial life was based on RNA only, which is known as *the RNA World hypothesis*. Since it was proven that RNA cannot only store and replicate genetic information but also catalyze chemical reactions, the RNA World hypothesis is quite fascinating (see Chap. 6). It has been

claimed, however, that abiotic synthesis of RNA is much more difficult than that of proteins. The synthesis of proteins requires two steps: (1) synthesis of amino acids and (2) polymerization of amino acids. On the other hand, the synthesis of RNAs requires five steps: (1) synthesis of nucleic acid bases, (2) synthesis of ribose, (3) synthesis of nucleosides from bases and ribose, (4) synthesis of nucleotides from nucleoside and phosphate, and (5) polymerization of nucleotides. Among these steps, the abiotic synthesis of nucleoside is the most difficult.

A great number of abiotic synthesis experiments have been conducted, most of which are based on the stepwise scenario of chemical evolution, as shown in Fig. 4.2. It could be said that each step is possible *in vitro*, but usually high concentrations of the starting materials are required to obtain the products. Powner et al. (2009) proposed a novel pathway of nucleotide synthesis in “prebiotically plausible” conditions (Fig. 4.6). Here they did not use nucleic acid bases and ribose as starting materials, but instead used high concentrations of starting molecules without any possible inhibitors and changed the reaction conditions of each reaction step (see Chap. 5 in detail).

Generally speaking, the prebiotic formation of oligopeptides is easier than that of oligonucleotides, since amino acids could be abiotically synthesized more easily than nucleotides. A great number of experiments have been conducted to condense amino acids via heating (Fox and Harada 1960), the use of clays as catalysts (Bujdák

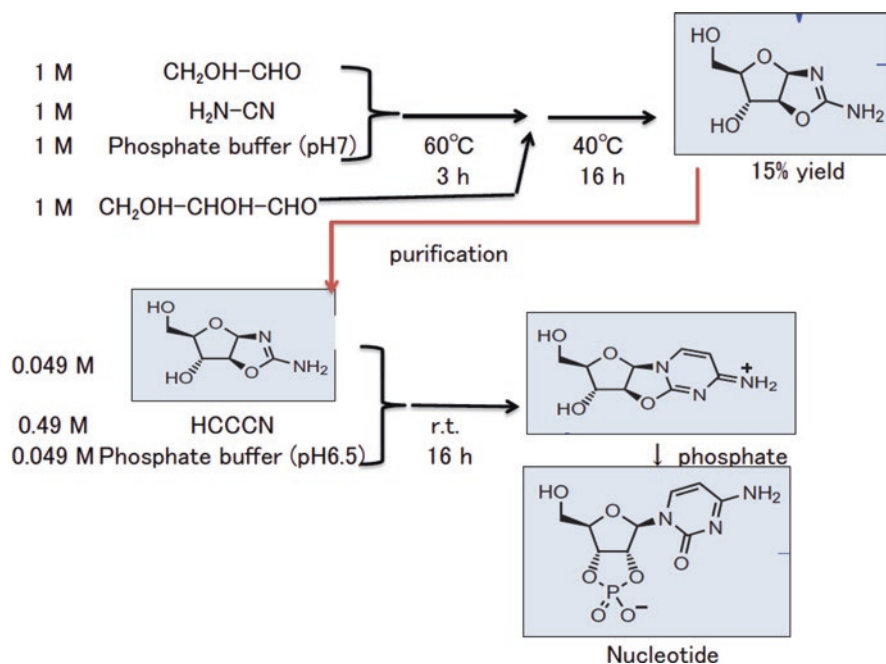


Fig. 4.6 Abiotic synthesis of nucleotides. (Based on the supplementary information from Powner et al. 2009)

and Rode 1999), and many other techniques. It was shown that oligopeptides could be formed from free amino acids. Amino acids are bifunctional molecules, such that it would be possible to synthesize long oligomers. It should be noted, however, that not only amino acids but also carboxylic acids, which are monofunctional molecules, were formed together with amino acids in many prebiotic synthesis experiments (Miller 1955). If such monofunctional molecules and amino acids coexisted, the elongation of oligopeptides would be terminated.

One of the largest differences between the abiotic world and the biotic world is that the former contains many more kinds of molecules than the latter. Schmitt-Kopplin et al. (2010) reported that extracts from the Murchison meteorite contained a large variety of molecules having more than 14,000 elemental compositions, which was estimated by comprehensive analysis using electrospray ionization-Fourier transform ion cyclotron resonance mass spectrometry (ESI-FTICR-MS). It is of importance to examine whether stepwise reactions toward the formation of biotic molecules could occur in systems containing a wide variety of abiotically available molecules.

4.5 Future Prospects

4.5.1 *Nobel Insights from Planetary Exploration*

Explorations to the moon and other planets started in 1959, and probes have visited all planets of the solar system, a number of their satellites, dwarf planets (Ceres and Pluto), asteroids, and comets. Among them, the Voyager and Cassini missions to the Saturnian moons (Titan and Enceladus) have given us a better understanding of prebiotic chemistry.

Titan is the largest satellite of Saturn, having a dense atmosphere predominantly composed of nitrogen and methane (Coustenis and Raulin 2015, see also Chap. 26). Astrobiologists have been highly interested in the chemistry of Titan's atmosphere due to the implications it may have in relation to the prebiotic chemistry that could have occurred in a slightly reducing primitive Earth atmosphere. The Voyager mission found the presence of various hydrocarbons and nitriles together with an orange haze in Titan's upper atmosphere; the haze was made of complex organic aerosols.

Sagan and Khare (1979) synthesized complex organic aerosols by plasma discharges in a simulated Titan atmosphere and named them *tholins*. They also found that amino acids were formed after the hydrolysis of *tholins* (Khare et al. 1986). A large number of experiments have been conducted to simulate possible organic reactions in Titan's atmosphere (Coll et al. 1998; Kobayashi et al. 2017). Most of them simulated possible reactions in Titan's upper atmosphere using energetic electrons and ultraviolet light, but it was shown that tholins containing amino acid precursors could be formed in dense gas mixtures in Titan's troposphere by way of such energy supplies as cosmic rays (Taniuchi et al. 2013).

The formation of complex amino acid precursors (*tholins*) by plasma discharge and particle irradiation in a simulated Titan atmosphere did not seem to result from a step-by-step process like the Strecker synthesis. Here instead, large molecular weight organics were directly formed by flash and quench type mechanisms. Such kinds of pathways should be considered for prebiotic reactions in the primitive Earth atmosphere and interstellar media.

It is not easy to presume the actual primitive Earth conditions, because they are lost on the present Earth. We are now getting more and more information on the organic reactions occurring in extraterrestrial bodies by way of planetary exploration. We will therefore be able to use these bodies as natural chemical evolution laboratories.

4.5.2 *Space Experiments*

In laboratory experiments simulating prebiotic conditions, a single energy has been usually used to examine the possible formation and alteration of organic compounds. Currently, however, we can perform space experiments by utilizing satellites and space stations, where actual space environments are available to examine prebiotic chemistry. For example, EXPOSE-E, EXPOSE-R, and EXPOSE-R2 were conducted from 2008 to 2016 by using European and Russian exposure facilities on the International Space Station (ISS), where a number of chemical and biological samples were exposed to space environments (Cottin 2015). The Tanpopo mission started in 2015 on the Exposure Facility of the Japanese Experiment Module (Kibo), which includes the exposure of microbial and chemical samples and a collection of dusts flying at high speed by using ultralow density silica aerogel (Kawaguchi et al. 2016).

In space, it would be possible to utilize the full solar spectrum and cosmic radiation at the same time, which would make possible to examine possible organic reactions in space that may have occurred on a primitive planet before the formation of the ozone layer. Such experiments might be done to examine the possible synergy of plural energies in prebiotic chemistry.

References

- Abe Y, Matsui T (1986a) Evolution of an impact-induced atmosphere and magma ocean on the accreting Earth. *Nature* 319:303–305
- Abe Y, Matsui T (1986b) Impact-induced atmospheres and ocean on Earth and Venus. *Nature* 322:526–528
- Abelson PHI (1966) Chemical events on the primitive Earth. *Proc Natl Acad Sci U S A* 55:1365–1372

- Barbier B, Bertrand M, Boillot F, Chabin A, Chaput D, Henin O, Brack A (1998) Extraterrestrial amino acids to the primitive Earth. Exposure experiments in Earth orbit. *Biol Sci Space* 12:92–95
- Bar-Nun A, Chang S (1968) Photochemical reactions of water and carbon monoxide in Earth's primitive atmosphere. *J Geophys Res* 88:6662–6672
- Bar-Nun A, Bar-Nun N, Bauer SH, Sagan C (1970) Shock synthesis of amino acids in simulated primitive environments. *Science* 168:470–473
- Bernstein MP, Dworkin JP, Sandford SA, Cooper GW, Allamandola LJ (2002) Racemic amino acids from the ultraviolet photolysis of interstellar ice analogues. *Nature* 416:401–403
- Bonner WA (1991) The origin and amplification of biomolecular chirality. *Orig Life Evol Biosph* 21:59–111
- Briggs R, Ertem G, Ferris JP, Greenberg JM, McCain PJ, Mendoza-Gomez CX, Schutte W (1992) *Orig Life Evol Biosph* 22:287–307
- Bujdák J, Rode BM (1999) The effect of clay structure on peptide bond formation catalysis. *J Mol Catal A* 144:129–136
- Burkov VI, Goncharova LA, Gusev GA, Kobayashi K, Moiseenko EV, Poluhina NG, Saito T, Tsarev VA, Xu J, Zhang G (2008) First results of the RAMBAS experiment on investigation of the radiation mechanism of chiral influence. *Orig Life Evol Biosph* 38:155–163
- Butlerow A (1861) Formation of a sugar-like substance by synthesis. *C R* 53:145–147
- Catling DC, Kasting JF (2017) Atmospheric evolution on inhabited and lifeless worlds. Cambridge University Press, Cambridge, UK, pp 231–237
- Chyba C, Sagan C (1991) Electrical energy sources for organic synthesis on the early earth. *Orig Life Evol Biosph* 21:3–17
- Chyba C, Sagan C (1992) Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature* 355:125–132
- Cody GD, Heying E, Alexander CMO, Nittler LR, Kilcoyne ALD, Sandford SA, Stroud RM (2011) Establishing a molecular relationship between chondritic and cometary organic solids. *Proc Natl Acad Sci U S A* 108:19171–19176
- Coll P, Coscia D, Gazeau M-C, Guez L, Raulin F (1998) Review and latest results of laboratory investigations of Titan's aerosols. *Orig Life Evol Biosph* 28:195–213
- Corliss JB, Dymond J, Gordon LI, Edmond JM, von Herzen RP, Ballard RD, Green K, Williams D, Bainbridge A, Crance K, van Andel TH (1979) Submarine thermal springs on the Galapagos rift. *Science* 203:1073–1083
- Cottin H (2015) Expose. In: Gargaud M, Irvine WM (eds) *Encyclopedia of astrobiology*, 2nd edn. Springer, Berlin, pp 812–814
- Coustenis A, Raulin F (2015) Titan. In: Gargaud M, Irvine WM (eds) *Encyclopedia of astrobiology*, 2nd edn. Springer, Berlin, pp 2506–2523
- Cronin JR, Pizzarello S (1997) Enantiomeric excesses in meteoritic amino acids. *Science* 275:951–955
- De Marcellus P, Meinert C, Nuevo M, Filippi J-J, Danger G, Deboffe D, Nahon L, d'Hendecourt LLS, Meierhenrich UJ (2011) *Astrophys J Lett* 727:L27 (6pp)
- Elsila JE, Glavin DP, Dworkin JP (2009) Cometary glycine detected in samples returned by stardust. *Meteor Planet Sci* 44:1323–1330
- Elsila J, Aponte JC, Blackmond DG, Burton AS, Dworkin JP, Glavin DP (2016) Meteoritic amino acids: diversity in compositions reflects parent body histories. *ACS Cent Sci* 2:370–379
- Ferris JP, Chen CT (1975) Chemical evolution XXVI. Photochemistry of methane, nitrogen and water mixtures as a model for the atmosphere of the primitive Earth. *J Am Chem Soc* 97:2962–2967
- Ferris JP, Sanchez RA, Orgel LE (1968) Studies in prebiotic synthesis. 3. Synthesis of pyrimidines from cyanoacetylene and cyanate. *J Mol Biol* 33:693–704
- Fox SW, Harada K (1960) The thermal copolymerization of amino acids common to protein. *J Am Chem Soc* 82:3745–3751

- Fukue T, Tamura M, Kandori R, Kusakabe N, Hough JH, Bailey J, Whittet DCB, Lucas PW, Nakajima Y, Hashimoto J (2010) Extended high circular polarization in the Orion massive star forming regions: implications for the origin of homochirality in the solar system. *Orig Life Evol Biosph* 40:335–346
- Fuller WD, Sanchez RA, Orgel LE (1972) Studies in prebiotic synthesis. VI Synthesis of purine nucleosides. *J Mol Biol* 67:25–33
- Garrison WM, Morrison DC, Hamilton JG, Benson A, Calvin M (1951) Reduction of carbon dioxide in aqueous solutions by ionizing radiation. *Science* 114:416–418
- Gilbert W (1984) The RNA world. *Nature* 319:618
- Greenberg JM, Li A (1997) Silicate core-organic refractory mantle particles as interstellar dust and as aggregated in comets and stellar disks. *Adv Space Res* 19:981–990
- Greenberg JM, Mendoza-Gomez CX (1993) Interstellar dust evolution: a reservoir of prebiotic molecules. In: Greenberg JM, Mendoza-Gomez CX, Pirronello V (eds) *The chemistry of life's origins*. Kluwer, Dordrecht, pp 1–32
- Gusev GA, Kobayashi K, Moiseenko EV, Poluhina NG, Saito T, Ye T, Tsarev VA, Xu J, Zhang G (2008) Results of the second stage of the investigation of the radiation mechanism of chiral influence (RAMBAS-2 experiment). *Orig Life Evol Biosph* 38:509–515
- Haldane, Maynard Smith J (ed) (1929) *The origin of life, rationalist annual*. In: *On being the right size and other essays*. Oxford University Press, Oxford. 1991
- Harada K, Fox SW (1960) The thermal copolymerization of aspartic acid and glutamic acid. *Arch Biochem Biophys* 86:274–280
- Harada K, Fox SW (1964) Thermal synthesis of natural amino-acids from a postulated primitive terrestrial atmosphere. *Nature* 210:335–336
- Imai E, Honda H, Hatori K, Brack A, Matsuno K (1999) Elongation of oligopeptides in a simulated submarine hydrothermal system. *Science* 283:831–833
- Inoue Y (1992) Asymmetric photochemical reactions in solution. *Chem Rev* 92:741–770
- Kasamatsu T, Kaneko T, Saito T, Kobayashi K (1997) Formation of organic compounds in simulated interstellar media with high energy particles. *Bull Chem Soc Jpn* 70:1021–1026
- Kasting JM (1990) Bolide impacts and the oxidation state of carbon in the Earth's early atmosphere. *Orig Life* 20:199–231
- Kawaguchi Y, Yokobori S, Hashimoto H, Yano H, Tabata M, Kawai H, Yamagishi A (2016) Investigation of the interplanetary transfer of microbes in the Tanpopo Mission at the Exposed Facility of the International Space Station. *Astrobiology* 16:363–376
- Kebukawa Y, Kilcoyne ALD, Cody GD (2013) Exploring the potential formation of organic solids in chondrites and comets through polymerization of interstellar formaldehyde. *Astrophys J* 771:19 (12pp)
- Kebukawa Y, Chan QHS, Tachibana S, Kobayashi K, Zolensky ME (2017) One-pot synthesis of amino acid precursors with insoluble organic matter in planetesimals with aqueous activity. *Sci Adv* 3:e1602093
- Khare BN, Sagan C, Ogino H, Nagy B, Er C, Schram KH, Arakawa ET (1986) Amino acids derived from Titan tholins. *Icarus* 68:176–184
- Kissel J, Krueger FR (1987) The organic component in dust from comet Halley as measured by PUMA mass spectrometer. *Nature* 326:755–760
- Kobayashi K, Ponnampereuma C (1985a) Trace elements and chemical evolution. *Orig Life* 16:41–55
- Kobayashi K, Ponnampereuma C (1985b) Trace elements in chemical evolution. II: synthesis of amino acids under simulated primitive Earth conditions in the presence of trace elements. *Orig Life* 16:57–67
- Kobayashi K, Saito T (2000) Energetics for chemical evolution on the primitive Earth. In: Akaboshi M, Fujii N, Navarro-Gonzalez R (eds) *The role of radiation in the origin and evolution of life*. Kyoto University Press, Kyoto, pp 25–37

- Kobayashi K, Kasamatsu T, Kaneko T, Koike J, Oshima T, Saito T, Yamamoto T, Yanagawa H (1995) Formation of amino acid precursors in cometary ice environments by cosmic radiation. *Adv Space Res* 16:21–26
- Kobayashi K, Kaneko T, Saito T, Oshima T (1998) Amino acid formation in gas mixtures by particle irradiation. *Orig Life Evol Biosph* 28:155–165
- Kobayashi K, Kaneko T, Takahashi J, Takano Y, Yoshida S (2010) High-molecular-weight complex organics in interstellar space and their relevance to origins of life. In: Basiuk VA (ed) *Astrobiology: emergence, search and detection of life*. American Scientific Publishers, Stevenson Ranch, pp 175–186
- Kobayashi K, Geppert WD, Carrasco N, Holm NG, Mousis O, Palumbo ME, Waite JH, Watanabe N, Ziurys LM (2017) Laboratory studies on methane and its relationship to prebiotic chemistry. *Astrobiology* 17:786–812
- Krueger K, Grabowski PJ, Zeug AJ, Sands J, Gottschling DE, Cech TR (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31:147–157
- Kurihara H, Yabuta H, Kaneko T, Obayashi Y, Takano Y, Kobayashi K (2012) Characterization of organic aggregates formed by heating products of simulated primitive Earth atmosphere experiment. *Chem Lett* 41:441–443
- Kvenvolden K, Lawless J, Pering K, Peterson E, Flores J, Ponnampereuma C, Kaplan IR, Moore C (1970) Evidence for extraterrestrial amino-acids and hydrocarbons in the Murchison meteorite. *Nature* 228:923–926
- Levy M, Miller SL, Oro J (1999) Production of guanine from NH_4CN polymerization. *J Mol Evol* 49:165–168
- Lohrmann R, Orgel LE (1971) Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science* 171:490–494
- Miller SL (1953) A production of amino acids under possible primitive earth conditions. *Science* 117:528–529
- Miller SL (1955) Production of some organic compounds under possible primitive earth conditions. *J Am Chem Soc* 77:2351–2361
- Miller SL, Orgel LE (1974) *The origin of life on the Earth*. Prentice-Hall, Englewood Cliffs
- Miller SL, Urey HC (1959) Organic compound synthesis on the primitive earth. *Science* 130:245–251
- Munoz Caro GM, Meerhenrich UJ, Schutte WA, Barbier B, Segavia AA, Rosenbauer H, Thiemann WHP, Brack A, Greenberg JM (2002) Amino acids from ultraviolet irradiation of interstellar ice analogues. *Nature* 416:403–406
- Oparin AI (1953) *The origin of life on Earth*, 3rd edn (trans: Synge A). Academic, New York
- Oro J (1960) Synthesis of adenine from ammonium cyanide. *Biochim Biophys Res Commun* 2:407–412
- Pace NR (1991) Origin of life – facing up to the physical setting. *Cell* 65:531–533
- Pizzarello S (2006) The chemistry of life's origin: a carbonaceous meteorite perspective. *Acc Chem Res* 39:231–237
- Ponnampereuma C, Lemmon RM, Mariner R, Calvin M (1963) Formation of adenine by electron irradiation of methane, ammonia and water. *Proc Natl Acad Sci USA* 49:737–740
- Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459:239–242
- Ring D, Wolman Y, Friedmann N, Miller SL (1972) Prebiotic synthesis of hydrophobic and protein amino acids. *Proc Natl Acad Sci USA* 69:765–768
- Sagan C, Khare BN (1971) Long wavelength ultraviolet photoproduction of amino acids on the primitive Earth. *Science* 173:417–420
- Sagan C, Khare BN (1979) Tholins: organic chemistry of interstellar grains and gas. *Nature* 277:102–107
- Sawai H (1976) Catalysis of internucleotide bond formation by divalent metal ions. *J Am Chem Soc* 98:7037–7039

- Schlesinger G, Miller SL (1983) Prebiotic synthesis in atmospheres containing CH₄, CO, and CO₂. I. Amino acids. *J Mol Evol* 19:376–382
- Schmitt-Kopplin P, Gabelica Z, Gougeon RD, Fekete A, Kanawati B, Harir M, Gebefuegi I, Eckel G, Herkorn N (2010) High molecular diversity of extraterrestrial organic matter in Murchison meteorite revealed 40 years after its fall. *Proc Natl Acad Sci U S A* 107:2763–2768
- Takano Y, Ohashi A, Kaneko T, Kobayashi K (2004) Abiotic synthesis of high-molecular-weight organics containing amino acid precursors from inorganic gas mixture of carbon nonoxide, ammonia and water by 3 MeV proton irradiation. *Appl Phys Lett* 84:1410–1412
- Takano Y, Takahashi J, Kaneko T, Marumo K, Kobayashi K (2007) Asymmetric synthesis of amino acid precursors in interstellar complex organics by circularly polarized light. *Earth Planet Sci Lett* 254:106–114
- Taniuchi T, Takano Y, Kobayashi K (2013) Amino acid precursors from a simulated lower atmosphere of Titan: experiments of cosmic ray energy source with ¹³C- and ¹⁸O-stable isotope probing mass spectrometry. *Anal Sci* 29:777–785
- Urey HC (1952) On the early chemical history of the Earth and the origin of life. *Proc Natl Acad Sci U S A* 38:351–363
- Yamagishi A, Yano H, Okudaira KK, Yokobori S, Tabata M, Kawai H, Yamashita M, Hashimoto H, Naraoka H, Mita H (2009) TANPOPO: astrobiology exposure and micrometeoroid capture experiments. *Trans Jpn Soc Aeronaut Space Sci Space Tech* 7(ists26):Tk_49–Tk_55
- Yanagawa H, Kobayashi K (1992) An experimental approach to chemical evolution in submarine hydrothermal systems. *Orig Life Evol Biosph* 22:147–159

Chapter 5

RNA Synthesis Before the Origin of Life



Yoshihiro Furukawa

Abstract Since RNA possesses both genetic information and catalytic abilities, it is a potential biopolymer that is thought to support primordial life before the present DNA-protein system. There were multiple sources that provided the building blocks of RNA and facilitated its synthesis on the prebiotic Earth. However, it remains unclear whether these sources provided a sufficient amount of building blocks. Moreover, several aspects of spontaneous construction of RNA, such as formation of glycosidic bonds and phosphoester bonds between building blocks, are yet to be elucidated. Nevertheless, recent studies have provided a number of insights into the spontaneous formation of RNA on the prebiotic Earth.

Keywords Prebiotic synthesis · RNA · Ribose · Nucleobase · Phosphate

5.1 Introduction

Some forms of biopolymers that can store genetic information are essential for life to realize Darwinian evolution. In the present life, the information contained in a gene is recorded as the sequence of nucleobases in DNA. This information in DNA is first transcribed to RNA; the information in the transcribed RNA eventually directs the synthesis of proteins, which actually work as biocatalysts. Therefore, RNA acts as a carrier in this conversion of genetic information from DNA to protein. Apart from RNA's role as the carrier of genetic information, its importance as a biocatalyst in the primordial life was also discussed nearly five decades ago. Later, supporting evidence including the discovery of ribozymes strengthened this notion and spread the RNA world hypothesis (see Chap. 6). Thus, spontaneous formation of RNA on the prebiotic Earth was a promising route to yield both catalytic and genetic biopolymers, simultaneously.

It was proposed that a biopolymer other than RNA may have supported primordial life; however, it was subsequently replaced by RNA. In this regard, several

Y. Furukawa (✉)

Department of Earth Science, Tohoku University, Sendai, Japan

e-mail: furukawa@tohoku.ac.jp

nucleic acid analogs have been proposed and investigated (e.g., Schneider and Benner 1990). Evolution of genetic and catalytic biopolymers has been discussed in details by Orgel (2004) and Engelhart and Hud (2010). Although the existence of a proto-RNA life has been discussed, it remains unclear which molecule could have fulfilled the role as a genetic material correctly work in the proto-RNA. On the other hand, it is amply clear that RNA plays essential roles in sustaining the present life forms transferring genetic information, and the RNA world might have preceded the DNA-protein world. To ensure the appearance of the RNA world, nucleotides had to be accumulated on primitive Earth.

However, it remains unclear how the primordial RNA was spontaneously formed on the prebiotic Earth. This process is not easy in terms of prebiotic chemistry and has been referred to as “The Molecular Biologist’s Dream” (Joyce and Orgel 1993). In this chapter, formation of oligoribonucleotides on the prebiotic Earth will be reviewed.

5.2 Availability of the RNA Components

5.2.1 Nucleobases

Adenine **2** (Fig. 5.1), guanine, cytosine, and uracil are the nucleobases constituting RNA. Several nucleobases have been found in carbonaceous meteorites. For example, the presence of uracil and xanthine in the Murchison meteorite has been reported (Martins et al. 2008). Similarly, purine nucleobases, adenine and guanine, have been found in several carbonaceous meteorites, particularly in CM2 chondrites (Callahan et al. 2011).

Formation of nucleobases also occurred on the prebiotic Earth. One of the processes that forms such organic compounds is the impacts of iron-bearing meteorites, which belong to the common type of meteorites collected on Earth to date. Meteorite impacts were far more frequent on the Hadean Earth than on the present Earth. Formation of pyrimidine nucleobases, cytosine and uracil, was demonstrated in simulated reactions of iron-bearing meteorite impacts on bicarbonate-ammonia-bearing ocean (Furukawa et al. 2015).

Formation of purine nucleobases upon heating the solutions containing hydrogen cyanide and ammonia was reported in the 1960s (Oró and Kimball 1961), whereas formation of pyrimidine nucleobases was described by cyanoacetaldehyde chemistry (Robertson and Miller 1995). If the prebiotic atmosphere was strongly reduced and rich in methane, ammonia, and hydrogen, hydrogen cyanide and cyanoacetaldehyde would have been formed by spark discharge in the atmosphere (Sanchez et al. 1966). However, doubt has been cast on the existence of such a strongly reduced atmosphere (Kasting 1993), and the primitive atmosphere will be discussed in another chapter.

5.2.2 Ribose

Another component of RNA **1** is ribose. Ribose is an aldopentose, which is the five-carbon sugar containing an aldehyde functional group. Pentoses including ribose are formed in the formose reaction in which condensation of formaldehyde under alkaline conditions results in formation of various carbohydrates (Butlerow 1860). However, aldopentoses including ribose are consumed with the progress of the formose reaction, since an aldehyde functional group in aldopentoses leads the molecule to further condensation reactions. Moreover, ribose is the least stable sugar among the four aldopentoses (i.e., ribose, arabinose, xylose, and lyxose) (Larralde et al. 1995). These two features of ribose appear to be the fundamental bottleneck for the possible accumulation of ribose on the prebiotic Earth.

In order to overcome these problems, effects of natural stabilizers have been proposed. The first stabilizer is borate, which can improve the stability of many sugars, in particular, ribose (Prieur 2001; Scorei and Cimpoiasu 2006; Furukawa et al. 2013). Benner and coworkers demonstrated that dissolved borate ion works as a stabilizer of ribose in simplified formose reactions (Ricardo et al. 2004; Kim et al. 2011). The second natural stabilizer is silicate. Silicate minerals are more abundant in nature, and dissolved silicate improves the stability of ribose (Lambert et al. 2010). The effects of these two stabilizers on the stability of aldopentoses have been compared quantitatively, and it has been shown that borate works on ribose more selectively and effectively than silicate does (Furukawa et al. 2013; Nitta et al. 2016). Although borate improves the stability of ribose, it also catalyzes the formation of branched pentoses (Kim et al. 2011). Geological evidences on Hadean Earth are not well preserved in contemporary rocks. The presence of borate minerals on the Hadean Earth is under hot debate (Grew et al. 2011). The mineral occurrences of early Archean rocks suggest the absence of borate minerals at that time, i.e., the absence of borate-containing fluid exceeding the saturation level of borate minerals. However, such geological evidences also suggest the presence of highly concentrated but undersaturated borate fluid at that time (Mishima et al. 2016). The availability of borate on the Hadean Earth is discussed in Furukawa and Kakegawa (2017), more extensively.

Ribose has never been detected in any astronomical sample yet, although a laboratory-based experiment demonstrated the formation of ribose and other sugars in ice analog of molecular clouds (Meinert et al. 2016). Sugar-related compounds, sugar alcohols and sugar acids, were found in Murchison and Murray meteorites (Cooper et al. 2001; Cooper and Rios 2016). Sugars have a chiral center; thus, they have enantiomeric forms (D- and L-forms). During chemical synthesis, the amounts of D- and L-sugars formed are the same, unless a chiral reagent is present in the reaction. Nevertheless, only D-ribose can be found in RNA. Cooper and Rios (2016) have found excess D-form of sugar acids in Murchison and Murray meteorites and suggested the possible origin of chirality of terrestrial life.

5.2.3 Phosphorus

Phosphorus is present as phosphate minerals such as hydroxyapatite, $\text{Ca}_5(\text{PO}_4)_3(\text{OH})$, in many igneous rocks. Actually, apatite crystals were found in zircon minerals formed 4.3 billion years ago on the Hadean Earth (Isozaki et al. 2017). Phosphorus is also found in meteorites in the form of phosphates and phosphides, such as schreibersite, $(\text{Fe,Ni})_3\text{P}$. Thus, phosphate was far more abundant than other RNA building blocks, ribose and nucleobases, on the prebiotic Earth.

5.3 Nucleoside Formation

There are two different approaches of nucleoside formation, namely, direct and indirect synthesis. Both reactions are dehydration forming a glycosidic bond. Thus, many attempts have been conducted simulating dry environments.

5.3.1 Direct Synthesis

Formation of a glycosidic bond between ribose and nucleobases has been attempted for some time, with limited success. Orgel and coworkers synthesized nucleosides from purine base and ribose. For example, adenosine **3** was synthesized from adenine **2** and ribose **1** and guanosine from guanine and ribose by heating and drying them with inorganic salts, although the yield of both adenosine and guanosine were only a few % (Fuller et al. 1972) (Fig. 5.1). This yield was further improved in a recent work using phosphorylated ribose (Kim and Benner 2017). Conversely, the reaction to form pyrimidine nucleoside with this method of direct synthesis remains unsuccessful.

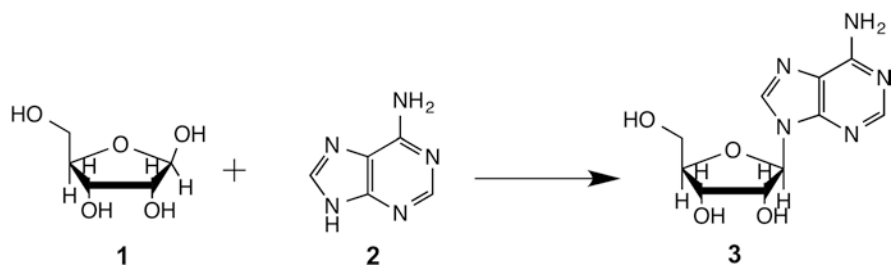


Fig. 5.1 Thermal dehydration using inorganic salts for glycosylation of adenine **2** by D-ribose **1** to form adenosine **3** (Fuller et al. 1972)

5.3.2 *Indirect Synthesis*

Indirect synthesis is the way to build up pentose and/or nucleobase moieties after making the glycosidic bond, instead of preparing pentose and nucleoside separately. Sanchez and Orgel have shown it is possible to form cytidine by reacting ribose, cyanimide, and cyanoacetylene in aqueous solutions (Sanchez and Orgel 1970). Sutherland and his coworkers have succeeded in forming a pyrimidine nucleoside without using ribose and nucleobase but by employing multistep reactions from cyanimide, cyanoacetylene, and aldehydes as starting materials (Powner et al. 2009). They further demonstrated the formation of a purine nucleoside derivative in a similar approach. More recently, formation of purine nucleoside in higher yields from ribose, guanidine, amide, nitrile, and ammonium cyanide has also been reported (Becker et al. 2016). However, these reactions employed cyanides, nitriles, and amides, which were rather scarce on the prebiotic surface of the Earth without a reducing atmosphere. Geochemical investigations in the future should explore the natural environment or events on the prebiotic Earth compatible to these sophisticated reactions.

5.4 Phosphorylation of Nucleosides to Form Nucleotides

Assuming the formation of a sufficient amount of nucleoside and phosphate, the formation of nucleotide combining these two substrates is not that obvious. There are two difficulties for the reaction: one is the low concentration of phosphate because of the low solubility in the presence of divalent cations, and the other is the presence of water, which promotes the hydrolysis of nucleotides.

5.4.1 *Presence of Abundant Water*

Phosphorylation of nucleosides is the reaction that results in the formation of phosphoester bonds. This reaction does not proceed effectively in aqueous solutions, since its reverse reaction, hydrolysis, predominates in water. In earlier work, phosphorylation of nucleosides was tested by simply heating and drying nucleoside solution with soluble phosphate (Ponnampertuma and Mack 1965). Lohrmann and Orgel (1971) have reported improved nucleotide yields using urea and ammonium chloride when heating and drying nucleoside solution with soluble phosphate and even with an insoluble phosphate, hydroxyapatite. More recently, nucleoside phosphorylation was demonstrated in a eutectic solvent of urea/ammonium formate/water with soluble and insoluble phosphates upon heating (Burcar et al. 2016), whereas Schoffstall (1976) reported phosphorylation of nucleosides using formamide, a nonaqueous solvent, though the presence of pure formamide as a solvent in the natural environment would have been possible only on the prebiotic Earth covered with reduced atmospheres, which is not consistent with the recent model.

5.4.2 Presence of Insoluble Phosphate

Phosphorus is present in many igneous rocks in the form of phosphates such as calcium phosphate mineral, hydroxyapatite. However, the solubility of these phosphate minerals is quite low.

A possible phosphorus material overcoming this issue is the reactive phosphorus in meteorites. Schreibersite is an iron phosphide mineral found in meteorites. This reduced phosphorus gets oxidized as soon as it reaches the surface of Earth and is retained in several oxidized forms before reaching the final oxidation state, phosphate. Pasek and coworkers analyzed the possibility of phosphorylation reaction using these reduced forms of phosphorus, such as phosphite. They observed that compared to phosphate, the reduced phosphorus species were more efficient in phosphorylating organic compounds, including nucleosides (Pasek and Lauretta 2005; Pasek et al. 2013; Gull et al. 2015).

Another possible prebiotic reaction circumvents the solubility problem is the involvement of borate phosphate mineral, lüneburgite, $\text{Mg}_3\text{B}_2(\text{PO}_4)_2(\text{OH})_6 \cdot 6\text{H}_2\text{O}$, containing both boron and phosphate. Kim et al. (2016) showed that up to 33% of adenosine **3** is phosphorylated to form adenosine 5'-monophosphate **5** selectively, when adenosine is subjected to phosphorylation by heating and drying with lüneburgite (Fig. 5.2).

5.5 Polymerization of Nucleotides

Polymerization of nucleotides is the final step in the abiotic RNA formation. This reaction also competes with its back reaction, the hydrolysis of the phosphoester bond. In the 1970s researchers tried to manage this problem using nucleoside cyclic phosphate **6**, which is more reactive than noncyclic phosphate because of the stress in its structure (Fig. 5.3). Orgel and coworkers have succeeded in forming more than six-monomer-long adenosine nucleotide from adenosine 2',3'-cyclic monophosphate by drying its alkaline solutions with simple catalyst such as aliphatic diamines (Verlander et al. 1973). Another possible reaction is the activation of nucleotides with imidazole. Imidazole derivative of nucleotides **8** spontaneously oligomerize, in particular with montmorillonite clay mineral, elongating 10-mer to 40-mer oligonucleotide (Ferris et al. 1996) (Fig. 5.4). In this reaction, imidazole derivatives of nucleotides are prepared in pure organic solvent with imidazole. However, geological settings that were compatible with the preparation of the imidazole derivatives of nucleotides on the prebiotic Earth are not found yet.

More recently, attempts have been made to form oligomers without activation. Rajamani et al. (2008) reported the formation of up to 100-mer oligonucleotides of adenosine monophosphate and uridine monophosphate by dehydration-rehydration of the 5'-monophosphate with phospholipids. However, subsequently, they characterized the product oligomers in details and showed that oligomers of purine

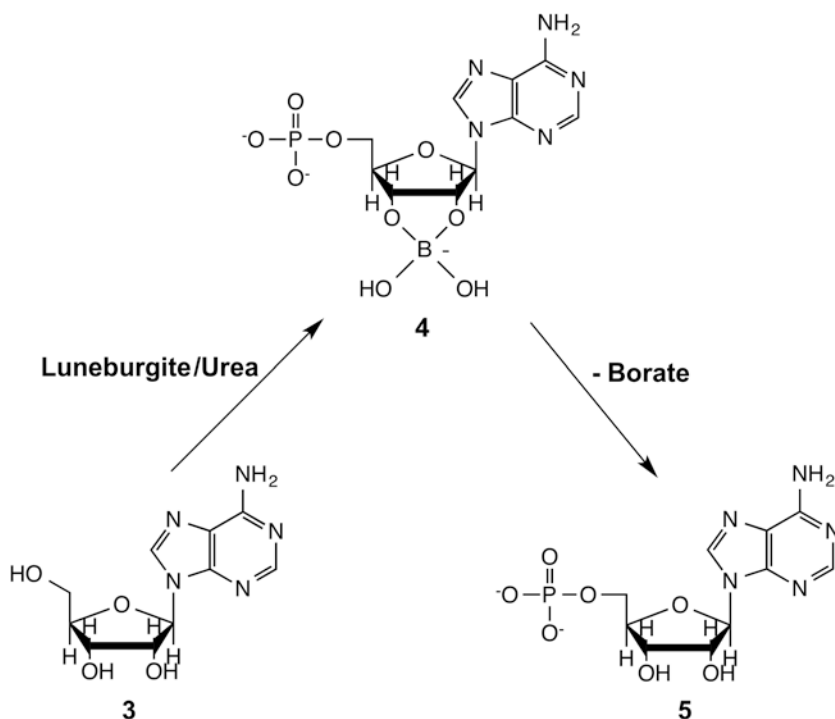
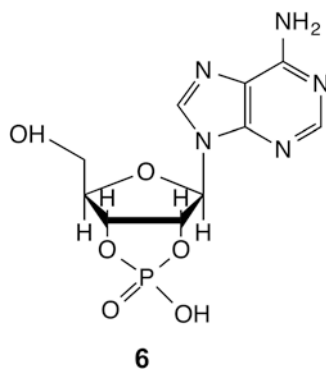


Fig. 5.2 Phosphorylation of adenosine **3** with lüneburgite, $\text{Mg}_3\text{B}_2(\text{PO}_4)_2(\text{OH})_6 \cdot 6(\text{H}_2\text{O})$. Adenosine combines with borate at the 2'- and 3'-hydroxyl group. Subsequently, phosphorylation selectively occurs at the 5'-hydroxyl group **4**. Borate is released by simple acidification of the solution, and adenosine 5'-phosphate **5** is selectively formed (Kim et al. 2016)

Fig. 5.3 Adenosine 2',3'-cyclic phosphate



nucleotide lost most of their base during the reaction, while the base loss was not substantial for the oligomers of pyrimidine nucleotide (Mungi and Rajamani 2015). Formation of long oligonucleotides (>100-mer) of adenosine monophosphate and guanosine monophosphate from adenosine 3',5'-cyclic monophosphate and

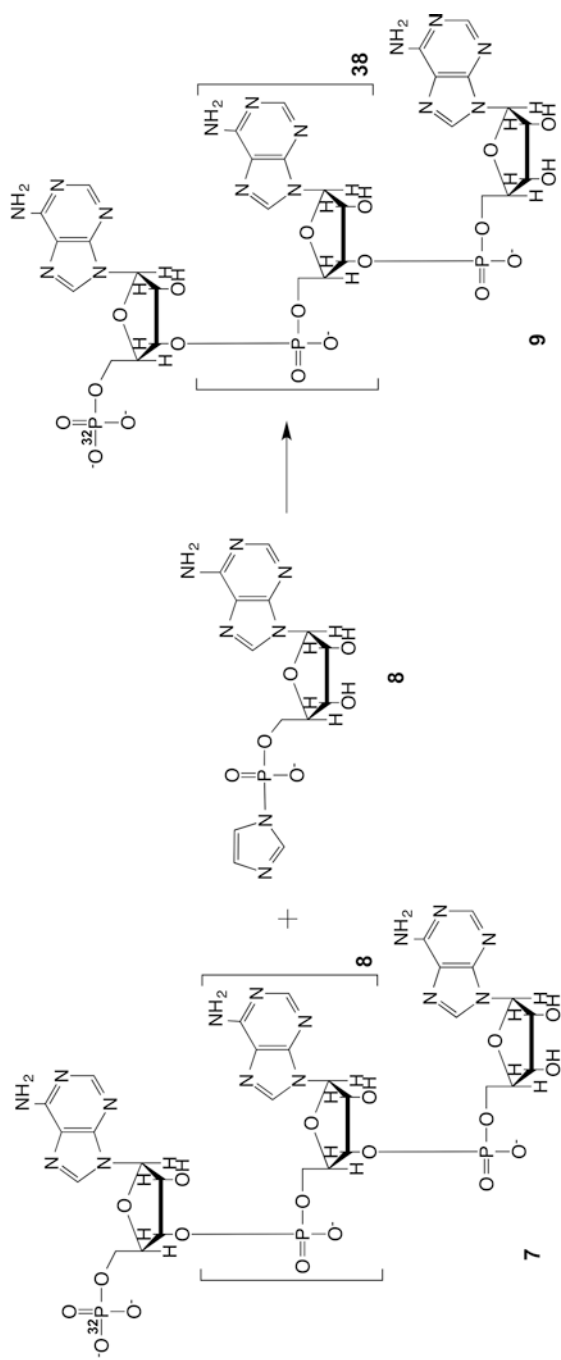


Fig. 5.4 Elongation of oligonucleotide with imidazole-activated monomer. Decamer of adenosine 5'-monophosphate **7** is elongated with imidazole-activated adenosine 5'-monophosphate **8** to form 40-mer of adenosine 5'-monophosphate (Ferris et al. 1996)

guanosine 3',5'-cyclic monophosphate, respectively, in water was also reported (Costanzo et al. 2009). However, this reaction has not been replicated consistently (Morasch et al. 2014). Instead, formation of long oligomers of guanosine monophosphate from guanosine 3',5'-cyclic monophosphate during a simple drying process was reported (Morasch et al. 2014). Nevertheless, spontaneous oligomerization of nucleotides under plausible prebiotic conditions is still one of the most challenging steps in primordial RNA formation.

5.6 Hadean Geological Settings Potentially Compatible to the RNA Formation

Many geological settings have been proposed for different steps in chemical evolution. The steps shown here for the formation of RNA require the concentration of reactants and the dehydration reaction. They would not have been achieved in seawater of open ocean, since reactants were too diluted. Hadean evaporitic environments have been discussed as a suitable geological setting for ribose formation, since such evaporitic environments promote concentration and dehydration (Fig. 5.5) (Furukawa and Kakegawa 2017). Such environments might have been compatible also for subsequent reactions of RNA formation such as nucleoside and nucleotide formation and even oligomerization of nucleotide to form primordial RNA, since such reactions are also dehydration and require the concentration of reactants and many mineral catalysts were available around such evaporitic environments close to Hadean juvenile crust.

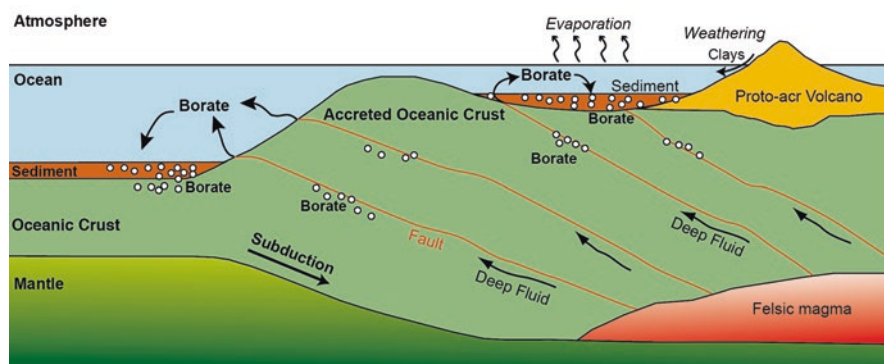


Fig. 5.5 Proto-arc model for the early Archean and Hadean and the proposed locations of Hadean evaporitic environments. White dots represent boron-rich clay. *TTG* tonalite–trondhjemite–granite suite (Furukawa and Kakegawa 2017)

5.7 Conclusions

Spontaneous formation of RNA on the prebiotic Earth is yet “Molecular Biologist’s Dream.” The availability of RNA building blocks from known sources is limited. Further, many steps in the synthesis of RNA remain to be understood. Nonetheless, recent progress in the field of prebiotic chemistry is steadily filling these gaps in the “Molecular Biologist’s Dream.” Future advances and interdisciplinary investigations in prebiotic chemistry, Hadean geochemistry, and RNA molecular biology would further enrich our understanding of the spontaneous formation of RNA on the prebiotic Earth.

Acknowledgment The author appreciates S. A. Benner and H-J. Kim for discussion on prebiotic chemistry and T. Kakegawa for discussion on Hadean geology. The author wishes to acknowledge JSPS KAKENHI (15K13588 and 15H03752) for financial support.

References

- Becker S, Thoma I, Deutsch A, Gehrke T, Mayer P, Zipse H, Carell T (2016) A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* 352(6287):833–836. <https://doi.org/10.1126/science.aad2808>
- Burcar B, Pasek M, Gull M, Cafferty BJ, Velasco F, Hud NV, Menor-Salván C (2016) Darwin’s warm little pond: a one-pot reaction for prebiotic phosphorylation and the mobilization of phosphate from minerals in a urea-based solvent. *Angew Chem Int Ed* 55(42):13249–13253. <https://doi.org/10.1002/anie.201606239>
- Butlerow A (1860) Ueber ein neues Methylenderivat. *Justus Liebigs Ann Chem* 115(3):322–327. <https://doi.org/10.1002/jlac.18601150325>
- Callahan MP, Smith KE, Cleaves HJ, Ruzicka J, Stern JC, Glavin DP, House CH, Dworkin JP (2011) Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *Proc Natl Acad Sci U S A* 108(34):13995–13998. <https://doi.org/10.1073/pnas.1106493108>
- Cooper G, Rios AC (2016) Enantiomer excesses of rare and common sugar derivatives in carbonaceous meteorites. *Proc Natl Acad Sci U S A* 113(24):E3322–E3331. <https://doi.org/10.1073/pnas.1603030113>
- Cooper G, Kimmich N, Belisle W, Sarinana J, Brabham K, Garrel L (2001) Carbonaceous meteorites as a source of sugar-related organic compounds for the early Earth. *Nature* 414(6866):879–883
- Costanzo G, Pino S, Ciciriello F, Di Mauro E (2009) Generation of long RNA chains in water. *J Biol Chem* 284(48):33206–33216. <https://doi.org/10.1074/jbc.M109.041905>
- Engelhart AE, Hud NV (2010) Primitive genetic polymers. *Cold Spring Harb Perspect Biol* 2(12):a002196. <https://doi.org/10.1101/cshperspect.a002196>
- Ferris JP, Hill AR, Liu R, Orgel LE (1996) Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* 381(6577):59–61
- Fuller WD, Orgel LE, Sanchez RA (1972) Studies in prebiotic synthesis 7: solid-state synthesis of purine nucleosides. *J Mol Evol* 1(3):249. <https://doi.org/10.1007/bf01660244>
- Furukawa Y, Kakegawa T (2017) Borate and the origin of RNA: a model for the precursors to life. *Elements* 13(4):261–265. <https://doi.org/10.2138/gselements.13.4.261>
- Furukawa Y, Horiuchi M, Kakegawa T (2013) Selective stabilization of ribose by borate. *Orig Life Evol Biosph* 43(4–5):353–361. <https://doi.org/10.1007/s11084-013-9350-5>

- Furukawa Y, Nakazawa H, Sekine T, Kobayashi T, Kakegawa T (2015) Nucleobase and amino acid formation through impacts of meteorites on the early ocean. *Earth Planet Sci Lett* 429:216–222. <https://doi.org/10.1016/j.epsl.2015.07.049>
- Grew ES, Bada JL, Hazen RM (2011) Borate minerals and origin of the RNA world. *Orig Life Evol Biosph* 41(4):307–316. <https://doi.org/10.1007/s11084-010-9233-y>
- Gull M, Mojica MA, Fernández FM, Gaul DA, Orlando TM, Liotta CL, Pasek MA (2015) Nucleoside phosphorylation by the mineral schreibersite. *Sci Rep* 5:17198. <https://doi.org/10.1038/srep17198>. <https://www.nature.com/articles/srep17198#supplementary-information>
- Isozaki Y, Yamamoto S, Sakata S, Obayashi H, Hirata T, Obori K-i, Maebayashi T, Takeshima S, Ebisuzaki T, Maruyama S (2017) High-reliability zircon separation for hunting the oldest material on Earth: an automatic zircon separator with image-processing/microtweezers-manipulating system and double-step dating. *Geosci Front*. <https://doi.org/10.1016/j.gsf.2017.04.010>
- Joyce GF, Orgel LE (1993) Prospects for understanding the origin of the RNA world. In: Gesteland RF, Atkins JF (eds) *The RNA world*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 1–25
- Kasting JF (1993) Earth's early atmosphere. *Science* 259(5097):920–926
- Kim H-J, Benner SA (2017) Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. *Proc Natl Acad Sci* 114(43):11315–11320. <https://doi.org/10.1073/pnas.1710778114>
- Kim H-J, Ricardo A, Illangkoon HI, Kim MJ, Carrigan MA, Frye F, Benner SA (2011) Synthesis of carbohydrates in mineral-guided prebiotic cycles. *J Am Chem Soc* 133(24):9457–9468. <https://doi.org/10.1021/ja201769f>
- Kim H-J, Furukawa Y, Kakegawa T, Bitá A, Scorei R, Benner SA (2016) Evaporite borate-containing mineral ensembles make phosphate available and regiospecifically phosphorylate ribonucleosides: borate as a multifaceted problem solver in prebiotic chemistry. *Angew Chem Int Ed* 55(51):15816–15820. <https://doi.org/10.1002/anie.201608001>
- Lambert JB, Gurusamy-Thangavelu SA, Ma KBA (2010) The silicate-mediated formose reaction: bottom-up synthesis of sugar silicates. *Science* 327(5968):984–986. <https://doi.org/10.1126/science.1182669>
- Larralde R, Robertson MP, Miller SL (1995) Rates of decomposition of ribose and other sugars: implications for chemical evolution. *Proc Natl Acad Sci U S A* 92(18):8158–8160. <https://doi.org/10.1073/pnas.92.18.8158>
- Lohrmann R, Orgel LE (1971) Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science* 171(3970):490. <https://doi.org/10.1126/science.171.3970.490>
- Martins Z, Botta O, Fogel ML, Sephton MA, Glavin DP, Watson JS, Dworkin JP, Schwartz AW, Ehrenfreund P (2008) Extraterrestrial nucleobases in the Murchison meteorite. *Earth Planet Sci Lett* 270(1–2):130–136. <https://doi.org/10.1016/j.epsl.2008.03.026>
- Meinert C, Myrgorodska I, de Marcellus P, Buhse T, Nahon L, Hoffmann SV, d'Hendecourt LLS, Meierhenrich UJ (2016) Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. *Science* 352(6282):208–212. <https://doi.org/10.1126/science.aad8137>
- Mishima S, Ohtomo Y, Kakegawa T (2016) Occurrence of tourmaline in metasedimentary rocks of the isua supracrustal belt, Greenland: implications for ribose stabilization in hadean marine sediments. *Orig Life Evol Biosph* 46(2):247–271. <https://doi.org/10.1007/s11084-015-9474-x>
- Morasch M, Mast CB, Langer JK, Schilcher P, Braun D (2014) Dry polymerization of 3',5'-cyclic GMP to long strands of RNA. *Chembiochem* 15(6):879–883. <https://doi.org/10.1002/cbic.201300773>
- Mungi CV, Rajamani S (2015) Characterization of RNA-like oligomers from lipid-assisted non-enzymatic synthesis: implications for origin of informational molecules on early Earth. *Life* 5(1):65–84. <https://doi.org/10.3390/life5010065>
- Nitta S, Furukawa Y, Kakegawa T (2016) Effects of silicate, phosphate, and calcium on the stability of aldopentoses. *Orig Life Evol Biosph* 46(2–3):189–202. <https://doi.org/10.1007/s11084-015-9472-z>

- Orgel LE (2004) Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol* 39(2):99–123. <https://doi.org/10.1080/10409230490460765>
- Oró J, Kimball AP (1961) Synthesis of purines under possible primitive earth conditions. I. Adenine from hydrogen cyanide. *Arch Biochem Biophys* 94(2):217–227. [https://doi.org/10.1016/0003-9861\(61\)90033-9](https://doi.org/10.1016/0003-9861(61)90033-9)
- Pasek MA, Lauretta DS (2005) Aqueous corrosion of phosphide minerals from iron meteorites: a highly reactive source of prebiotic phosphorus on the surface of the early Earth. *Astrobiology* 5(4):515–535
- Pasek MA, Hammmeijer JP, Buick R, Gull M, Atlas Z (2013) Evidence for reactive reduced phosphorus species in the early Archean ocean. *Proc Natl Acad Sci U S A* 110(25):10089–10094. <https://doi.org/10.1073/pnas.1303904110>
- Ponnamperuma C, Mack R (1965) Nucleotide synthesis under possible primitive earth conditions. *Science* 148(3674):1221–1223. <https://doi.org/10.1126/science.148.3674.1221>
- Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459(7244):239–242 http://www.nature.com/nature/journal/v459/n7244/supinfo/nature08013_S1.html
- Prieur BE (2001) Étude de l'activité prébiotique potentielle de l'acide borique. *C R Acad Sci Paris Chim Chem* 4(8–9):667–670. [https://doi.org/10.1016/s1387-1609\(01\)01266-x](https://doi.org/10.1016/s1387-1609(01)01266-x)
- Rajamani S, 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(1):57–74. <https://doi.org/10.1007/s11084-007-9113-2>
- Ricardo A, Carrigan MA, Olcott AN, Benner SA (2004) Borate minerals stabilize ribose. *Science* 303(5655):196–196
- Robertson MP, Miller SL (1995) An efficient prebiotic synthesis of cytosine and uracil. *Nature* 375(6534):772–774. <https://doi.org/10.1038/375772a0>
- Sanchez RA, Orgel LE (1970) Studies in prebiotic synthesis: V. Synthesis and photo-anomerization of pyrimidine nucleosides. *J Mol Biol* 47(3):531–543. [https://doi.org/10.1016/0022-2836\(70\)90320-7](https://doi.org/10.1016/0022-2836(70)90320-7)
- Sanchez RA, Ferris JP, Orgel LE (1966) Cyanoacetylene in prebiotic synthesis. *Science* 154(3750):784. <https://doi.org/10.1126/science.154.3750.784>
- Schneider KC, Benner SA (1990) Oligonucleotides containing flexible nucleoside analogs. *J Am Chem Soc* 112(1):453–455. <https://doi.org/10.1021/ja00157a073>
- Schoffstall AM (1976) Prebiotic phosphorylation of nucleosides in formamide. *Orig Life Evol Biosph* 7(4):399–412. <https://doi.org/10.1007/bf00927935>
- Scorei R, Cimpoiu VM (2006) Boron enhances the thermostability of carbohydrates. *Orig Life Evol Biosph* 36(1):1–11. <https://doi.org/10.1007/s11084-005-0562-1>
- Verlander MS, Lohrmann R, Orgel LE (1973) Catalysts for the self-polymerization of adenosine cyclic 2',3'-phosphate. *J Mol Evol* 2(4):303–316. <https://doi.org/10.1007/bf01654098>

Part III
History of Life Reveiled from Bilology

Chapter 6

RNA World



Shotaro Ayukawa, Toshihiko Enomoto, and Daisuke Kiga

Abstract In the RNA world, which is a hypothetical idea to explain the origin of life, RNA molecules were considered to have roles in both information storage and as a catalyst. This chapter reviews RNA world studies from its birth to recent advancements. Natural ribozymes and coenzymes containing nucleotide moieties support the hypothesis. For maintenance and evolution of the RNA world, ribozymes that have self-replicating activity had to emerge. Although such a ribozyme has not been discovered in the natural world, in vitro evolution experiments have created ribozymes that have replicative ability, essentially. After the emergence of the replicative ribozyme, ribozymes might have gradually improved its activity by incorporating other biomolecules especially peptides, which could be synthesized spontaneously in the prebiotic world. The RNA-protein (RNP) world may have emerged through the interaction between the RNA world and peptide/protein world. Proteins with higher enzymatic activities could have appeared through Darwinian evolution in the RNP world, and the peptide/protein must have replaced the role of ribozymes as catalysts. Further interactions with other molecular worlds such as the lipid or metabolic worlds accelerated the evolution of the self-replicating system. Finally, DNA, which is chemically more stable than RNA, has taken over the role as the storage of genetic information.

Keywords RNA · Ribozyme · Peptide · Self-replication · In vitro evolution

S. Ayukawa (✉)

Waseda Research Institute for Science and Engineering, Waseda University, Tokyo, Japan
e-mail: ayukawa@aoni.waseda.jp

T. Enomoto · D. Kiga (✉)

Department of Electrical Engineering and Bioscience, Waseda University, Tokyo, Japan
e-mail: toshihiko@asagi.waseda.jp; kiga@waseda.jp

6.1 Introduction

RNA world is a hypothetical idea that there was a period, in the evolutionary history of life, in which RNA molecules served as both information storage carriers and catalysts. Though the original hypothesis had assumed that the RNA world was composed of RNA only (Gilbert 1986), recent studies have expanded the hypothesis to include separately evolved biomolecules such as peptides which cooperatively worked with RNA (Krishnamurthy 2017).

In the present world, namely, the DNA-protein world, genetic information is stored in DNA, while catalytic activities are performed by protein enzymes (Fig. 6.1). In this system, proteins are produced using genetic information stored in DNA. However, without the enzymatic activities of proteins, protein cannot be produced from the genetic information. This situation suggests the existence of chicken-and-egg paradox: which came first, proteins or DNA. This paradox is solved by the RNA world hypothesis that assumes the existence of self-replicating RNA molecules with capacities for both the storage of genetic information and enzymatic activities. There is no doubt that RNA can store genetic information as well as DNA, considering the ability of RNA to hybridize with complementary strands. However, there was no indication prior to 1982 that RNA could also have enzymatic activity.

The discovery of RNA molecules with enzymatic activity or ribozymes triggered wide acceptance of the RNA world theory. Though the enzymatic activities of RNA molecules had been predicted in the 1960s by Woese (1967), Crick (1968), and Orgel (1968), biochemical evidence was absent. At the time, proteins were believed to be only biological molecules with catalytic activities. In the early 1980s, however,

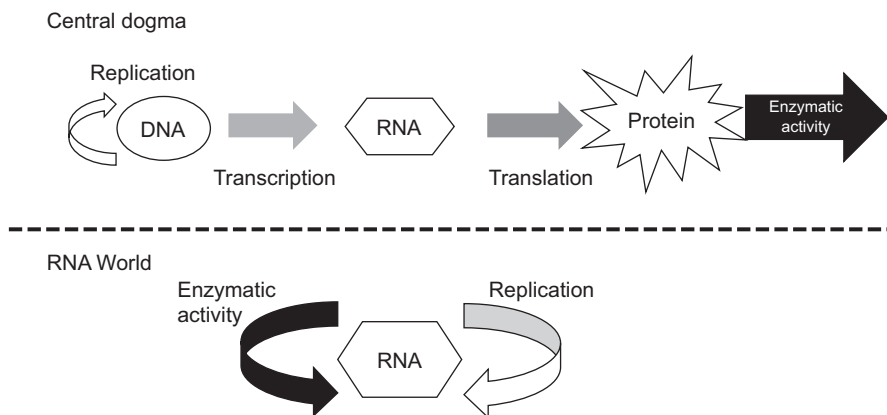


Fig. 6.1 Central dogma and the RNA world. Flow of genetic information from DNA to protein is called the central dogma. DNA stores genetic information. RNA serves as a mediator of DNA information. Proteins behave as functional molecules translated from genetic information. In the RNA world, on the other hand, RNA served as the storage of genetic information as well as the functional molecule

the Cech's (Kruger et al. 1982) and Altman's (Guerrier-Takada et al. 1983) groups independently discovered RNA molecules with catalytic activities inside living cells. Based on these facts, Walter Gilbert proposed the RNA world hypothesis in 1986 (Gilbert 1986). After the first discoveries of ribozymes, several other types of ribozymes were found (Prody et al. 1986; Hutchins et al. 1986; Buzayan et al. 1986; Michel et al. 1989; Kuo et al. 1988; Feldstein et al. 1989; Wu et al. 1989; Saville and Collins 1990; Nelson and Breaker 2017).

After the proposal of the RNA world hypothesis, the ribonucleotide-derived coenzymes working in modern organisms were recognized as molecular fossils of the RNA world. These coenzymes were possibly synthesized in the RNA world and could have been used for ribozymes in a variety of metabolisms to maintain the RNA world.

Though self-replicating RNA has not been identified in modern organisms, ribozymes that can catalyze RNA synthesis have been developed experimentally by several groups. Additionally, ribozymes harboring the potential to interact with other types of molecules were created. In this chapter, studies that support the RNA world hypothesis are reviewed, and the transition from the RNA world to the RNP and DNA-protein worlds is also discussed.

6.2 Life in the RNA World

In the RNA world, RNA molecules served as carriers of genetic information and as catalysts both for metabolism and self-replication. These ribozymes are thought to be generated through natural selection initiated from random ribonucleotide sequences. Interactions with other biomolecules such as peptides and lipids, which evolved separately from the RNA world in the prebiotic era, could have contributed to broaden the catalytic abilities of the ribozymes (Krishnamurthy 2017). These interactions might have led the primitive system to modern organisms through the RNP world.

6.3 Molecular Fossils of the RNA World: Ribozymes and Coenzymes

A ribozyme or RNA molecule that has enzymatic activity is a typical example of RNA world vestiges. One of the first discovered ribozymes was an intron of the rRNA precursor of *Tetrahymena thermophila*. The ribozyme had self-splicing activity, and it catalyzed its own excision from the rRNA precursor under the existence of guanosine and magnesium ions. The splicing is triggered by specific binding of guanosine on the group I intron. The 3'-OH group of the guanosine attacks to the phosphodiester bond at 5' exon. Concurrently, Sidney Altman found that at the

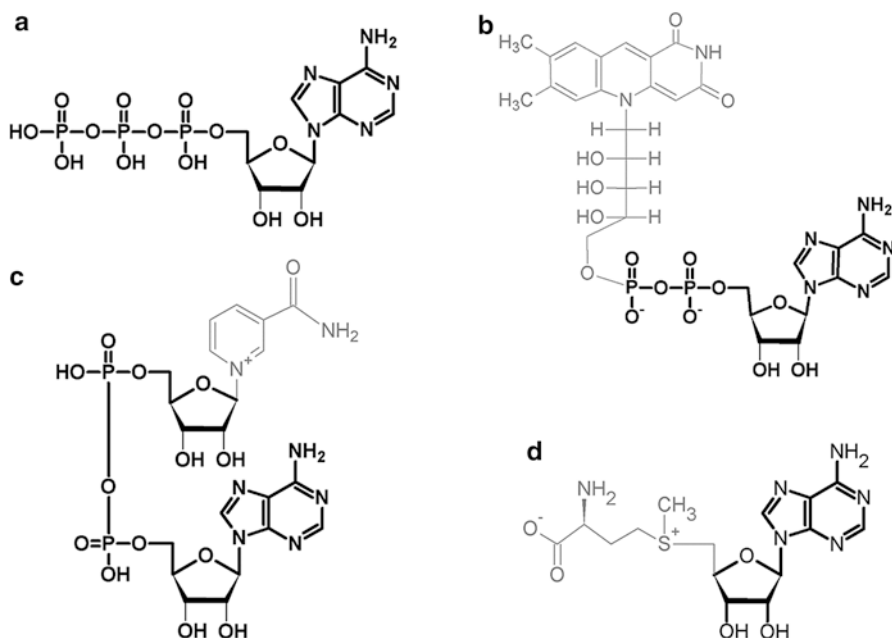


Fig. 6.2 Molecular fossils from the RNA world. Coenzymes often contain ribonucleotide moieties. RNA components in the coenzymes are colored in black. (a) Adenosine triphosphate (ATP), (b) flavin adenine dinucleotide (FAD), (c) nicotinamide adenine dinucleotide (NAD⁺), and (d) S-adenosylmethionine (SAM)

center of ribonuclease P (RNase P), which is the key enzyme of tRNA processing, is an RNA molecule (Guerrier-Takada et al. 1983). RNase P, which specifically cleaves precursors of tRNAs to generate mature tRNAs, had been known to be a complex of a protein and an RNA molecule. Altman's group discovered that the RNA molecule alone maintains this processing activity without the aid of the protein component.

A ribozyme also controls the peptidyl transferase activity of modern ribosomes. The ribosome, which is a complex of RNA molecules and proteins, serves as the site for translation in all types of cells. Peptide bonds between amino acids are formed to produce proteins by the peptidyl transferase reaction performed in the ribosome. Based on the X-ray crystal structure of the 50S subunit of the ribosome, there is no protein within 25 Å of the active site of the peptidyl transferase (Nissen et al. 2000). This result indicates that the peptidyl transferase activity of the ribosome is carried by RNA molecules. A main role of the proteins in the ribosome is thought to be the stabilization of the ribosome structure.

Derivatives of RNA molecules used as coenzymes in a variety of enzymatic reactions in modern cells also support the RNA world hypothesis (Nelson and Breaker 2017). Many coenzymes such as adenosine triphosphate (ATP) (Fig 6.2a), flavin adenine dinucleotide (FAD) (Fig 6.2b), nicotinamide adenine dinucleotide (NAD⁺)

(Fig. 6.2c), and S-adenosyl methionine (SAM) (Fig. 6.2d) are commonly used in all three domains of modern life. There is no doubt that ribozymes composed of only four kinds of nucleotides required a variety of coenzymes for their activities in the RNA world because even modern protein enzymes, with 20 amino acid building blocks, a number much larger than 4, require coenzymes. The small compounds derived from nucleotide metabolism in the RNA world were bound with the ribozymes to facilitate various reactions in the RNA world and remain as present coenzymes supporting contemporary proteinaceous enzymes (Fig. 6.2). Recently, specific cellular RNAs were shown to be linked with NAD⁺ at its 5'-terminus although the function of the coenzyme moiety has not been identified (Cahová et al. 2015).

6.4 Expansion of RNA Biomass: From Accumulation by Chemical Evolution to Replication by Macromolecular Catalyst

One of the main issues in the birth of the RNA world is the emergence of self-replicating RNA molecules. Gerald Joyce and Leslie Orgel called this event “molecular biologist’s dream.” Though a clear answer for the problem has not yet been obtained, a feasible scenario for the emergence of the self-replicator could be as follows (Fig. 6.3): during chemical evolution, firstly, short oligo-RNA strands were generated by random assembly of ribonucleotide molecules. From the library of oligo-RNA strands, a strand with RNA ligase activity has emerged. The emerged ligase ribozymes began to assemble the oligo-RNA strands around them, and longer RNA strands were generated. From the long RNA strands, the strands with RNA polymerase activity must have emerged. The first generation of the RNA polymerase ribozymes may not have had high fidelity; thus, a large number of mutant RNA strands might have been generated. From the mutant library, the RNA polymerase ribozyme with higher elongation activity and fidelity than the parent ribozyme must have emerged through random mutation and selection, namely, Darwinian evolution. The containment of RNA by lipid vesicle must have accelerated such evolution (Szostak et al. 2001). Such an RNA polymerase ribozyme became the self-replicator and flourished in the RNA world. This section introduces the studies that attempted to confirm possible self-replication in the RNA world. Though the remnant of self-replicase ribozymes has not been discovered in the present world, various types of ribozymes with RNA ligase or RNA polymerase activities that are necessary for self-replication in the RNA world have been developed experimentally. Moreover, theoretical studies started by Eigen set the limit that explained the possibility of the self-replication (Eigen 1977). Additionally, ribozymes that catalyze some metabolism producing nucleotide molecules are important for the expansion of RNA world biomass, and some of such ribozymes have indeed been developed through in vitro evolution (Unrau and Bartel 1998).

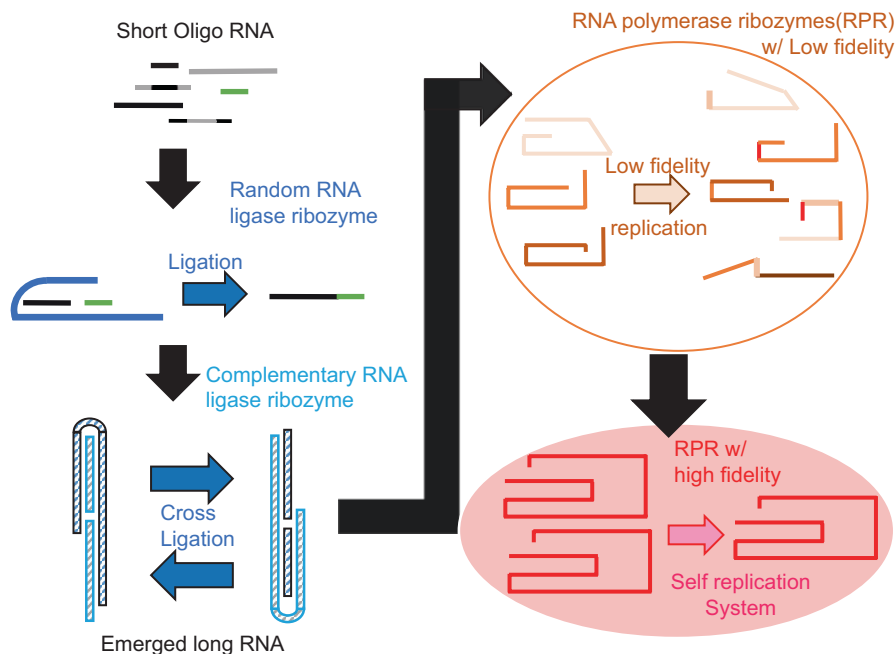


Fig. 6.3 Transition from oligo-RNA groups to the RNA world. Short oligo-RNAs were present in the prebiotic environment. The RNA ligase ribozyme that could randomly ligate oligo-RNAs emerged spontaneously. Next complementary ligase ribozymes could generate long RNA strands. As a result, the RNA polymerase ribozyme (RPR) appeared. The early-stage RPR was not able to function as self-replicator. However, RPR spontaneously acquired high fidelity and gained the ability to self-replicate. The RNA world started at that time so that RNA could maintain genetic information

In order to support the existence of the RNA world, the mechanism of self-replication of an RNA strand has to be understood. RNA replication can be achieved through two types of reactions: replication with ribozyme catalysis and that without ribozyme catalysis. Ribozyme-independent replication was shown to occur when activated nucleotides exist in a reaction mixture (Inoue and Orgel 1983; Orgel 2004). However, the fidelity of enzyme-independent replication is not enough to achieve the synthesis of an RNA strand that is long enough to have some catalytic activity. In the RNA world, a ribozyme that has an ability to replicate itself, or a RNA replicase ribozyme, must have existed.

In the early stages of ribozyme study, a group I intron of *Tetrahymena thermophila* that has self-splicing activity was found to catalyze elongation of RNA strands like an RNA polymerase (Been and Cech 1988). Though the group I intron cannot be called a RNA replicase ribozyme due to its inability to synthesize itself, its elongation activity can be part of the reactions that has to be catalyzed by the RNA replicase ribozyme.

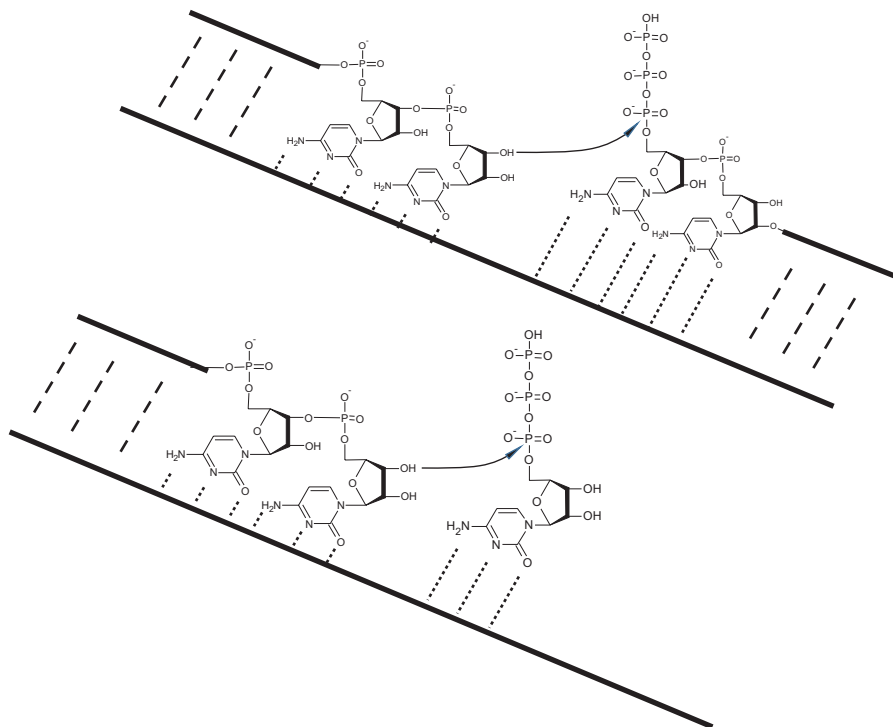


Fig. 6.4 Chemistry shared in ligation and polymerization. Both reactions involve the formation of a 3-5' phosphodiester bond between the 3'-terminus of the RNA strand and 5'-phosphate of the substrate. The difference between these reactions is the size of the substrates: short oligo-RNA for ligation and mononucleotide for polymerization

The existence of an RNA replicase ribozyme has been shown by studies that have developed ribozymes in a laboratory through *in vitro* evolution. The *in vitro* evolution is a method that mimics the process of natural selection to evolve RNA sequence toward a desired function. The sequences that meet the selection criteria can be obtained through iterative rounds of selection, mutagenesis, and amplification. With this *in vitro* evolution method, activity of the natural group I introns was improved (Green and Szostak 1992). Moreover, new ribozymes have been created from random sequence libraries, including the ribozymes with catalytic activities that do not exist in the present living world.

Such *in vitro* evolution methods have created ribozymes with partial characteristics of the RNA replicase. To accomplish self-replication by RNA molecules, the key reaction is the formation of a 3-5' phosphodiester bond between the 3'-terminus of the RNA strand and 5'-terminus of the substrate. Thus, RNA replicase ribozymes could be either ligases or polymerases, both of which share the same chemistry (Fig. 6.4). The RNA ligase ribozyme catalyzes template-directed assembly of the short oligo-RNA substrates to generate a long RNA strand, while

the RNA polymerase ribozyme catalyzes elongation of the RNA strand using one ribonucleotide at a time, as a substrate.

Several types of ligase ribozymes that catalyze template-directed formation of the 3–5' phosphodiester bond between two RNA strands were generated, in the early stages of *in vitro* evolution study. Bartel and Szostak were able to create a ligase ribozyme that can ligate two RNA strands (Bartel and Szostak 1993). The RNA ligase, named Class I ligase, was selected from the random RNA sequences, 220 nucleotides in length, by the *in vitro* evolution method. Class I ligase catalyzed the ligation of an exogenous fragment of RNA strands to its own 5' terminus. Following Class I ligase, Ellington's group showed that even a much shorter RNA strand, L1 ligase, isolated from random sequences 90 nucleotides in length, can have ligase activity (Robertson and Ellington 1999). Although the secondary structure of the L1 ligase is much simpler than that of the Class I ligase, the L1 ligase functioned to ligate two RNA strands to form a 3–5' phosphodiester bond. This result showed that a short RNA strand can be an RNA ligase. Since shorter RNA strands could be synthesized non-enzymatically in the prebiotic world more efficiently than longer ones, the existence of such a short artificial RNA ligase supports the spontaneous generation of an RNA replicase ribozyme.

Subsequently, Joyce and his coworkers have created a self-replication system composed of two types of RNA ligases, each of which catalyzed the assembly of the other one (Paul and Joyce 2002). They prepared the ribozyme which catalyzed the assembly of two RNA fragments, 13-mer and 48-mer RNAs, which were the components of the counterpart ligase. Though there is no evidence that such a ligase ribozyme pair could have existed in the prebiotic environment, similar pairs could be considered as a prototype of self-replication system in the RNA world.

Next, *in vitro* evolution was used for the creation of ribozymes with the activity of an RNA-dependent RNA polymerase that catalyzes the formation of phosphodiester bond using ribonucleotide triphosphates (NTPs) as substrate (Table 6.1). Bartel and his coworker succeeded in isolating an RNA polymerase ribozyme from the Class I ligase ribozyme (Eklund and Bartel 1996) (Table 6.1-1). The 98-mer RNA polymerase ribozyme had the ability to extend an RNA primer by six bases at most by assembling NTP. However, the ribozyme can only use a template strand containing a specific sequence because a part of the ribozyme strand is needed to hybridize to the specific sequence of the template strand to keep the template in the active center. Moreover, it took 4 days for the six-base extension. Bartel and his coworker pushed forward with the improvement of this RNA polymerase ribozyme to obtain the ribozyme that can extend an RNA strand by 14 bases in 1 day. Furthermore, the ribozyme could recognize a template strand regardless of its sequence (Johnston et al. 2001) (Table 6.1-2). Subsequently, using the ribozyme as the starting strand, Holliger and his coworker selected a new RNA polymerase ribozyme that achieved the extension of NTPs up to 95 bases (Wochner et al. 2011) (Table 6.1-3). Even with this achievement, its activity was insufficient to replicate the ribozyme itself, which is 198-nucleotide long. Later, the same group developed an RNA polymerase ribozyme that could extend an RNA strand up to 206 nucleotides, while the length of the ribozyme is 202-nucleotide long (Attwater et al. 2013)

Table 6.1 Abilities of RNA polymerase ribozymes

Name	Extension length [nt]	Length of ribozyme [nt]	Error rate	Condition	Ref
b1-233t (Table 6.1-1)	6	96	15×10^{-2}	60 mM MgCl ₂ 22 °C, 6 days	Ekland and Bartel (1996)
Round-18 ribozyme (Table 6.1-2)	14	189	3.3×10^{-2}	200 mM MgCl ₂ , 22 °C, 24 h	Johnston et al. (2001)
tC19 (Table 6.1-3)	95	198	2.7×10^{-2}	200 mM MgCl ₂ , 17 °C, 7 days	Wochner et al. (2011)
tC9Y (Table 6.1-4)	206	202	2.3×10^{-2}	200 mM MgCl ₂ , 17 °C, 7 days	Attwater et al. (2013)
tC9-4M (Table 6.1-5)	195	177	2.2×10^{-2}	10 mM MgCl ₂ + 6 μM K ₁₀ , 17 °C, 21 days K ₁₀ : lysine decapeptide	Tagami et al. (2017)

(Table 6.1-4). More recently, a 165-nucleotide long RNA polymerase ribozyme that could extend an RNA strand up to 198 nucleotides was developed (Tagami et al. 2017) (Table 6.1-5). Since these ribozymes can synthesize RNA strands that are longer than their own lengths, they are capable of replicating themselves if the ribozymes had enough fidelity.

Fidelity is also an important characteristic of the RNA replicating ribozyme because accumulated errors in replication interfere with the maintenance of genetic information in the RNA world (Table 6.1). The longer the sequence of the molecule, the more difficult it becomes to maintain the correct genetic information, because of increase in errors during replication. In other words, the fidelity of a self-replication limits the length of the molecule to be maintained.

Through the analysis that shows the requirements for maintenance of the amount of information during repeated cycles of replication, Eigen proposed an idea that information does not exist by itself, but is instead continuously maintained as long as the information is decoded to molecules with function (Eigen 1977). In his claim, furthermore, the continuous replication of the information molecules requires the molecules to have selective advantage compared with the relatively similar mutants that are produced by errors in the replication of the information. When the fidelity of the replicase is low, large amounts of mutant molecules are produced compared with the number of correctly replicated molecules. For the maintenance of the original information, thus, the correctly produced molecule has to have higher selective advantage than the mutant molecules. This analysis gave an upper limit of information content from the fidelity and the selective advantage; this limit is called Eigen limit. In summary, if an RNA polymerase ribozyme overcomes the Eigen limit, it can become the self-replicator. This limitation is also applied for a multiple gene system containing a replicase and other enzymes. For the establishment of such a complex system, a genome has to contain large information that is coded by a long sequence of the genome, and the replicase must have high fidelity of replication.

Not only nucleotide sequence but also structures and concentrations of other chemical compounds can be the information in the self-replication system containing RNA and other compounds such as peptides and lipids produced by the system.

6.5 The World After the RNA World

The RNA world is considered to have evolved into the present DNA-protein world, which is developed by the accumulation of interactions between many heterogeneous molecules such as DNA, protein, RNA, and fatty acids (Krishnamurthy 2017). While the RNA world appeared in the prebiotic era, several other types of molecular worlds have been proposed (Fig. 6.5). For example, nonenzyme- and non-templated-dependent polymerization of amino acids could result in accumulation of short oligopeptides. From the randomly accumulated peptides, longer chains of peptides with some catalytic activities could be generated. These peptides might accelerate the expansion of the peptide world. At this time, potential ribozymes that could catalyze peptide synthesis could make a small positive feedback system among RNAs and peptides. Indeed, the activities of some ribozymes can be increased through interaction with small peptides (Tagami et al. 2017). Similar to such a positive feedback mechanism, a lipid world consisting of accumulated fatty acids could also interact with the RNA world. Lipid layers were shown to be able to accelerate dehydration synthesis of activated nucleotides (Rajamani et al. 2008). On

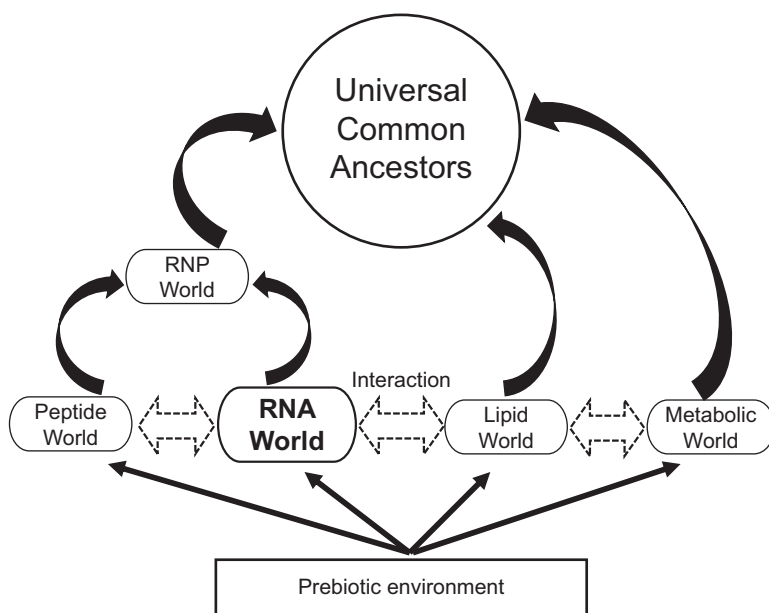


Fig. 6.5 Interaction of the worlds before common ancestors

the other side of the feedback loop, ribozyme reactions can affect liposomal properties (Chen et al. 2005). The emergence of these and other heterogeneous molecular worlds could be the basis for the transition of the RNA world to the modern world.

Before emergence of the present template-directed synthesis of proteins, it is thought that ribozymes that catalyze peptide synthesis were generated. Protein synthesis in present cells consists of two reactions: (1) synthesis of aminoacyl-tRNA by associating amino acids with corresponding tRNAs and (2) connection of a peptide on a tRNA and an amino acid on the other tRNA. Both tRNAs are aligned by the order of codons on the mRNA template strand. In present life, reaction (1) is catalyzed by aminoacyl-tRNA synthetase, and reaction (2) is catalyzed by ribosomal RNA in the ribosome. For protein synthesis to occur in the RNA world, ribozymes that catalyze these two reactions had to be generated.

Of the two important reactions for protein synthesis, aminoacyl-tRNA synthesis involves two steps, each of which was shown to be catalyzed by a ribozyme generated by *in vitro* evolution. In the first step, aminoacyl AMP, which is the amino acid connected to AMP, is produced by activating an amino acid with the energy of an ATP molecule. In the second step, the activated amino acid is transferred to a hydroxyl group on the 3'-terminal ribose of the tRNA. Yarus and his coworker have produced a ribozyme that catalyzes the second step of the reaction in 1995 (Illangasekare et al. 1995). This ribozyme had the capability to transfer a phenylalanine of phenylalanyl-AMP to its own 3' terminus. Furthermore, Suga and his coworker produced a ribozyme that transfers phenylalanine activated with the cyanomethyl group, a simpler leaving group, to the 3' terminus of tRNA (Saito et al. 2001). Later, regarding the first step, the ribozyme that catalyzes the aminoacyl-AMP synthesis was produced by Yarus and his coworker in 2001 (Kumar and Yarus 2001). The ribozyme had the capacity to synthesize an aminoacyl-AMP from an amino acid and an ATP. The generation of these ribozymes indicates that all of the present catalytic activities producing aminoacyl-tRNA could have occurred in the RNA world. Not only catalytic activity but also specificity is important for a genetic code. Creation of a series of amino acid-binding aptamers, the molecule that has binding ability, has shown that RNA can discriminate an amino acid from the other types of amino acids (Famulok 1994; Majerfeld and Yarus 1994; Yarus 2017).

Peptidyl transferase activity, another important reaction in protein synthesis, was also accomplished by ribozymes. In present life forms, peptidyl transferase activity is encompassed by the large rRNA in the ribosome. Cech and his coworker have shown that the peptidyl transferase activity can be accomplished with a simpler RNA molecule obtained by *in vitro* evolution (Zhang and Cech 1997). They have prepared RNA library covalently linked an amino acid at its 5' terminus for the selection. Peptidyl transferase activity of ribozymes in the library transferred an amino acid moiety of another molecule that mimicked the terminal structure of aminoacyl-tRNA to the other amino acids covalently linked to the ribozyme. This result showed that not only the generation of the aminoacyl-tRNA but also the peptide bond formation could be accomplished with ribozymes. Although further study is required to attain the codon-anticodon pairing-dependent peptidyl transferase activity, the transition from the RNA world to the RNP world is shown to be chemically feasible.

Along with the formation of the RNP world, heterogeneous interactions with other molecular species might have accelerated further evolution of the RNA-based living system. A variety of reactions and small molecules developed in the metabolic world could diversify the activities of ribozymes and proteins. Compartmentalization of these molecules by lipid membranes improved the efficiencies of the chemical reactions by increasing local concentrations of the substrates and enzymes. At some point in the transition, protein or RNA enzymes with the reverse transcriptase activity that synthesize DNA strands from RNA templates must have emerged. RNA has 2'-OH which is the critical group in the ribozyme catalytic reaction. RNA is chemically not stable because of the 2'-OH, which is not present in DNA. DNA, which is chemically more stable than RNA, took over the role as the storage of genetic information. With the double-stranded structure of the DNA, without the 2'-OH group, genetic information is more securely maintained.

6.6 Conclusion

In the RNA world, RNA molecules stored genetic information and had catalytic activity. The first proposed RNA world hypothesis assumed that the RNA world was composed of RNA molecules only, but recent studies have shown an advanced view that the RNA world was supported by other molecules such as peptides.

For the maintenance and evolution of the RNA world, ribozymes that have the catalytic activity to reproduce themselves had to emerge. Though such a ribozyme has not been discovered in the contemporary living organisms, the in vitro evolution method have succeeded in finding ribozymes with ligase or polymerase activity.

The RNA world is considered to be a stepping stone in the evolution to the present life world. In the prebiotic world, RNA molecules that gained self-replicating activities formed the RNA world. The RNA world might have gradually improved its activity by involving biomolecules especially peptides, which were synthesized spontaneously in the prebiotic world. Such improvement was accelerated when some RNA molecules gained the ability to synthesize peptides with specific sequences. This was the start of RNA-protein (RNP) world which later transitioned into the DNA-protein world. Though this scenario has not been fully proven experimentally, additional studies would clarify the details of the evolution of the RNA world.

References

- Attwater J, Wochner A, Holliger P (2013) In-ice evolution of RNA polymerase ribozyme activity. *Nat Chem* 5:1011–1018. <https://doi.org/10.1038/nchem.1781>
- Bartel DP, Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences. *Science* 261(5127):1411–1418. <https://doi.org/10.1126/science.7690155>
- Been MD, Cech TR (1988) RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the Tetrahymena ribozyme. *Science* 239(4846):1412–1416. <https://doi.org/10.1126/science.2450400>

- Buzayan JM, Hampel A, Bruening G (1986) Nucleotide sequence and newly formed phosphodiester bond of spontaneously ligated satellite tobacco ringspot virus RNA. *Nucleic Acids Res* 14(24):9729–9743. <https://doi.org/10.1093/nar/14.24.9729>
- Cahová H, Winz ML, Höfer K, Nübel G, Jäschke A (2015) NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* 519(7543):374–377. <https://doi.org/10.1038/nature14020>
- Chen IA, Salehi-Ashtiani K, Szostak JW (2005) RNA catalysis in model protocell vesicles. *J Am Chem Soc* 127(38):13213–13219. <https://doi.org/10.1021/ja051784p>
- Crick FH (1968) The origin of the genetic code. *J Mol Biol* 38(3):367–379. [https://doi.org/10.1016/0022-2836\(68\)90392-6](https://doi.org/10.1016/0022-2836(68)90392-6)
- Eigen M (1977) The hypercycle a principle of natural self-organization part a: emergence of the hypercycle. *Naturwissenschaften* 64:541–565
- Eklund EH, Bartel DP (1996) RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* 382(6589):373–376. <https://doi.org/10.1038/382373a0>
- Famulok M (1994) Molecular recognition of amino acids by RNA-aptamers: an L-citrulline binding RNA motif and its evolution into an L-arginine binder. *J Am Chem Soc* 116(5):1698–1706. <https://doi.org/10.1021/ja00084a010>
- Feldstein PA, Buzayan JM, Bruening G (1989) Two sequences participating in the autolytic processing of satellite tobacco ringspot virus complementary RNA. *Gene* 82(1):53–61. [https://doi.org/10.1016/0378-1119\(89\)90029-2](https://doi.org/10.1016/0378-1119(89)90029-2)
- Gilbert W (1986) Origin of life – the Rna world. *Nature* 319(6055):618–618. <https://doi.org/10.1038/319618a0>
- Green R, Szostak JW (1992) Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science* 258(5090):1910–1915. <https://doi.org/10.1126/science.1470913>
- Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35(3 Pt 2):849–857 doi:0092-8674(83)90117-4 [pii]
- Hutchins CJ, Rathjen PD, Forster AC, Symons RH (1986) Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res* 14(9):3627–3640. <https://doi.org/10.1093/nar/14.9.3627>
- Illangasekare M, Sanchez G, Nickles T, Yarus M (1995) Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* 267(5198):643–647. <https://doi.org/10.1126/science.7530860>
- Inoue T, Orgel LE (1983) A nonenzymatic RNA polymerase model. *Science* 219(4586):859–862. <https://doi.org/10.1126/science.6186026>
- Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* 292(5520):1319–1325. <https://doi.org/10.1126/science.1060786>
- Krishnamurthy R (2017) Giving rise to life: transition from prebiotic chemistry to protobiology. *Acc Chem Res* 50(3):455–459. <https://doi.org/10.1021/acs.accounts.6b00470>
- Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31(1):147–157. [https://doi.org/10.1016/0092-8674\(82\)90414-7](https://doi.org/10.1016/0092-8674(82)90414-7)
- Kumar RK, Yarus M (2001) RNA-catalyzed amino acid activation. *Biochemistry* 40(24):6998–7004. <https://doi.org/10.1021/bi010710x>
- Kuo MY, Sharmeen L, Dinter-Gottlieb G, Taylor J (1988) Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *J Virol* 62(12):4439–4444
- Majerfeld I, Yarus M (1994) An Rna pocket for an aliphatic hydrophobe. *Nat Struct Biol* 1(5):287–292. <https://doi.org/10.1038/nsb0594-287>
- Michel F, Umeson K, Ozeki H (1989) Comparative and functional anatomy of group II catalytic introns – a review. *Gene* 82(1):5–30. [https://doi.org/10.1016/0378-1119\(89\)90026-7](https://doi.org/10.1016/0378-1119(89)90026-7)
- Nelson JW, Breaker RR (2017) The lost language of the RNA world. *Sci Signal* 10(483). doi:<https://doi.org/10.1126/scisignal.aam8812>

- Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* 289(5481):920–930. <https://doi.org/10.1126/science.289.5481.920>
- Orgel LE (1968) Evolution of the genetic apparatus. *J Mol Biol* 38(3):381–393. [https://doi.org/10.1016/0022-2836\(68\)90393-8](https://doi.org/10.1016/0022-2836(68)90393-8)
- Orgel LE (2004) Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol* 39(2):99–123. <https://doi.org/10.1080/10409230490460765>
- Paul N, Joyce GF (2002) A self-replicating ligase ribozyme. *Proc Natl Acad Sci U S A* 99(20):12733–12740. <https://doi.org/10.1073/pnas.202471099>
- Prody GA, Bakos JT, Buzayan JM, Schneider IR, Bruening G (1986) Autolytic processing of dimeric plant virus satellite RNA. *Science* 231(4745):1577–1580. <https://doi.org/10.1126/science.231.4745.1577>
- Rajamani S, 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(1):57–74. <https://doi.org/10.1007/s11084-007-9113-2>
- Robertson MP, Ellington AD (1999) In vitro selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat Biotechnol* 17(1):62–66. <https://doi.org/10.1038/5236>
- Saito H, Kourouklis D, Suga H (2001) An in vitro evolved precursor tRNA with aminoacylation activity. *EMBO J* 20(7):1797–1806. <https://doi.org/10.1093/emboj/20.7.1797>
- Saville BJ, Collins RA (1990) A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora mitochondria*. *Cell* 61(4):685–696. [https://doi.org/10.1016/0092-8674\(90\)90480-3](https://doi.org/10.1016/0092-8674(90)90480-3)
- Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. *Nature* 409:387–390. <https://doi.org/10.1038/35053176>
- Tagami S, Attwater J, Holliger P (2017) Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function. *Nat Chem* 9:325–332. <https://doi.org/10.1038/nchem.2739>
- Unrau PJ, Bartel DP (1998) RNA-catalysed nucleotide synthesis. *Nature* 395(6699):260–263. <https://doi.org/10.1038/26193>
- Wochner A, Attwater J, Coulson A, Holliger P (2011) Ribozyme-catalyzed transcription of an active ribozyme. *Science* 332(6026):209–212. <https://doi.org/10.1126/science.1200752>
- Woese CR (1967) *The genetic code: the molecular basis for genetic expression*. Harper & Row, New York
- Wu HN, Lin YJ, Lin FP, Makino S, Chang MF, Lai MM (1989) Human hepatitis delta virus RNA subfragments contain an autocleavage activity. *Proc Natl Acad Sci U S A* 86(6):1831–1835
- Yarus M (2017) The genetic code and RNA-amino acid affinities. *Life (Basel)* 7(2):13. <https://doi.org/10.3390/life7020013>
- Zhang B, Cech TR (1997) Peptide bond formation by in vitro selected ribozymes. *Nature* 390(6655):96–100. <https://doi.org/10.1038/36375>

Chapter 7

The Common Ancestor of All Modern Life



Satoshi Akanuma

Abstract All modern organisms on Earth share a common mechanism for replication and expression of genetic material. Given the complexity of the genetic mechanism, it seems unlikely that the same construct developed independently in different organisms. Therefore, a reasonable hypothesis is that all modern organisms on Earth are descendants of a single common ancestral organism, and the common ancestor already had the basic genetic mechanism found in modern organisms. A phylogenetic tree that illustrates the evolutionary paths of organisms also shows that all existing organisms originate from a single root that is located between the last common archaeal and bacterial ancestors. Recently published articles on the universal ancestor suggest that it was an anaerobic autotroph dependent on H_2 and CO_2 from geochemical sources and surrounded by a cell membrane similar to those found in modern bacteria and eukaryotes. In contrast to conflicting conclusions of *in silico* studies on the environmental temperature of the universal ancestor, reconstruction of ancestral protein sequences and characterization of their properties *in vitro* suggest that the universal ancestor was a thermophile or hyperthermophile that thrived at a very high temperature. Future research may continue to revise these predictions of features associated with the universal ancestor.

Keywords Anaerobic autotroph · Ancestral sequence reconstruction · Cell membrane · Single ancestry · Thermophilicity

7.1 Introduction

In the *On the Origin of Species* (Darwin 1859), Darwin predicted that all modern organisms are descendants of a single common ancestor. The fact that all modern organisms share a significantly similar mechanism for replication and expression of genetic information supports the existence of the single ancestor. It should be noted that the universal common ancestor is not our oldest ancestor but rather the most

S. Akanuma (✉)

Faculty of Human Sciences, Waseda University, Tokorozawa, Saitama, Japan
e-mail: akanuma@waseda.jp

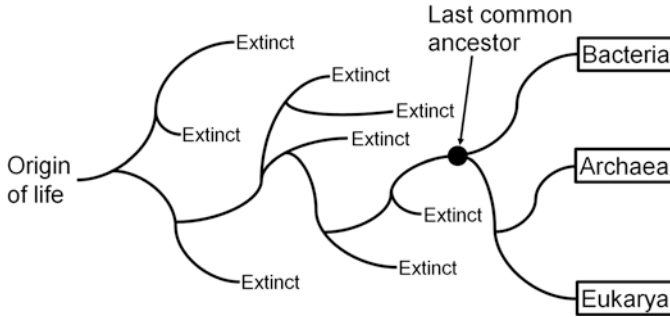


Fig. 7.1 Evolution before and after the last universal common ancestor. After the origin of life, diversification of life would have occurred for a few million years. Therefore, a variety of organisms likely existed at the time of the last universal common ancestor. However, most of these primitive organisms became extinct for various reasons, including meteorite impacts, and only the last universal common ancestor survived and further evolved

recently existing common ancestor of all modern organisms. Diversification of life would have occurred for a few million years after first appearing on Earth (the origin of life) ~4200 to 4000 million years ago (Mya) (Cornish-Bowden and Cardenas 2017). However, most of these primitive organisms likely became extinct for various reasons, including meteorite impacts, leaving only the universal common ancestor (Fig. 7.1) (Nisbet and Sleep 2001). To distinguish the first life and the universal common ancestor, the latter is often called the “last universal common ancestor,” where “last” means “most recent.” A molecular clock analysis suggests that the last universal common ancestor lived about 3800 Mya (Feng et al. 1997) which is compatible with geochemical surveys that suggest life had already emerged at ~3400 to 4300 Mya (Schopf 1993; Rosing 1999; Shen et al. 2001; Wacey et al. 2006; Ueno et al. 2006; Bell et al. 2015; Nutman et al. 2016; Dodd et al. 2017). This chapter summarizes current knowledge about the last universal common ancestor.

7.2 Is All Modern Life a Descendant of a Single Ancestor?

Not all biologists support the single common ancestor hypothesis. Woese (Woese and Fox 1977) thought that organisms prior to speciation into archaea, bacteria, and eukarya were far simpler than the modern prokaryote. Moreover, he thought that the universal common ancestor was a population of fast-evolving entities that mated and exchanged genetic material. He called these primitive entities progenotes, in which genotype and phenotype had not yet been completely linked as in modern organisms. Kandler (1995) also did not support the theory of single ancestry. He proposed a pre-cell theory where organisms before speciation formed a population of pre-cells that had metabolic and self-replication abilities and mutually interchanged their genetic information within the population. Doolittle (1999) thought

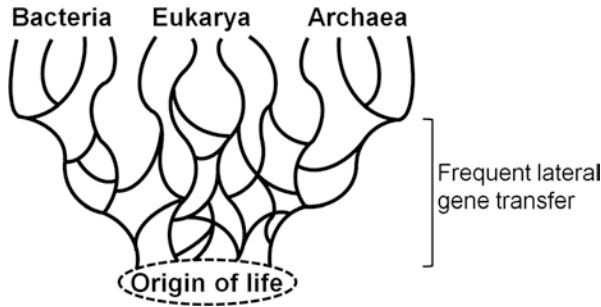


Fig. 7.2 A tree, or net, of life proposed by Doolittle (1999). He thought that lateral gene transfer occurred more frequently in early stages of evolution than at present and therefore that the early history of life should be represented as a net

that lateral gene transfer occurred more frequently in early stages of evolution than at present, and therefore he proposed a tree of life as represented in Fig. 7.2.

However, all modern life on Earth shares a common basic genetic mechanism. It is based on DNA as a genetic molecule that contains four different bases; proteins are used as functional molecules and consist of the same 20 L-amino acids encoded by the universal genetic code table; basic components of transcription and translation are generally the same among all life. It would be difficult to explain how the same genetic mechanism was established many times independently, and it therefore seems more reasonable to assume that all modern organisms on Earth are descendants of a single common ancestral organism and that the common ancestor already had the rigid genetic mechanism found in organisms today (Yamagishi et al. 1998).

Theobald (2010) used a formal statistical test to answer the question of whether all modern life on Earth is a descendant from a single common ancestor. He tested the universal common ancestry hypothesis by constructing evolutionary trees of 23 universally conserved proteins. Then, he compared the likelihood values of monophyly with those for different ancestry hypotheses and found stronger support for monophyly of all modern organisms than for other evolutionary models, even when horizontal gene transfers were considered. Therefore, the common ancestry model is currently better supported than other evolutionary models.

7.3 What to Name the Last Universal Common Ancestor?

The last universal common ancestor has been referred to by many different names. As mentioned above, Woese called it progenote, whereas other authors referred to it as cenancestor (Doolittle and Brown 1994) or *Commonote* (Yamagishi et al. 1998). Although it has most commonly been referred to as LUCA, I hereafter refer to it as *Commonote commonote* or *C. commonote* (Akanuma et al. 2015). Unlike progenote, *C. commonote* is defined as a species with a single genotype that had a stable cell membrane.

7.4 Placement of *C. commonote* on a Tree of Life

The evolutionary pathway of modern life can be visualized as a phylogenetic tree. A genetic tree based on ribosomal RNA has been most referenced as a species tree (Woese 1987; Woese et al. 1990) and separates modern organisms into three domains. Some phylogenetic and phylogenomic studies also support the three-domain hypothesis (Harris et al. 2003; Ciccarelli et al. 2006; Yutin et al. 2008; Rinke et al. 2013) (Fig. 7.3a). However, other studies instead suggest two domains in a tree of life (Rivera and Lake 1992; Cox et al. 2008; Williams et al. 2013; Raymann et al. 2015; Furukawa et al. 2017) (Fig. 7.3b).

Placement of *C. commonote* on a tree generally requires inclusion of two paralogous proteins that duplicated before *C. commonote* emerged (Akanuma et al. 2013b). Such composite trees were independently reported from studies of elongation factor (Iwabe et al. 1989) and H⁺-ATPase (Gogarten et al. 1989). In both trees, *C. commonote* was placed between the archaeal and bacterial common ancestors. In contrast, Cavalier-Smith (2002, 2006a, b, 2010) and Lake and coworker (2008, 2009) suggested that *C. commonote* was a type of bacteria because the roots were contained within the *Bacteria* domain in their trees. However, most composite trees that were subsequently built support the position of the root between the archaeal and bacterial common ancestors (Miyazaki et al. 2001; Brown and Doolittle 1995; Fournier et al. 2011). Thus, the current consensus is that *C. commonote* is located at the branch that connects the common ancestors of *Archaea* and *Bacteria* (Fig. 7.3).

7.5 Physiology of *C. commonote*

Efforts to predict the gene content of *C. commonote* have focused on genes that are universally distributed among modern organisms. However, this approach may not be valid because frequent occurrence of horizontal transfer and loss of genes during evolution prevents accurate predictions of the gene composition of *C. commonote* (Fournier et al. 2015; Groussin et al. 2015). Therefore, Martin and coworkers (Weiss

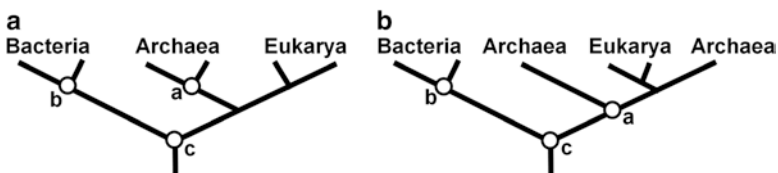


Fig. 7.3 Three domain (a) or two domain (b) hypotheses of life. In this chapter, the last universal common ancestor is referred to as *Commonote commonote* or *C. commonote*, which is defined as a species with a single genotype that had a stable cell membrane. Similarly, the last common ancestor of *Archaea* and of *Bacteria* is referred to as *C. archaea* and *C. bacteria*, respectively (Akanuma et al. 2015). The positions of *C. archaea*, *C. bacteria*, and *C. commonote* are indicated with a, b, and c, respectively

et al. 2016) removed from their analysis any proteins that might have experienced lateral gene transfer and traced back remaining proteins to *C. commonote*. They identified 355 genes that plausibly existed in *C. commonote* and used this gene set to reconstruct *C. commonote*'s metabolic pathways. Inclusion of a rotator–stator ATP synthase subunit and H⁺/Na⁺ antiporters in the gene set suggests that, for energy metabolism, *C. commonote* used ion gradients possibly generated by a geochemical mechanism. The gene content also suggests that *C. commonote* was an anaerobic autotroph that largely relied on H₂ and CO₂ that may have been abundant in its environment. This inference is compatible with the predicted environment of the primitive biosphere where the amount of oxygen was negligible.

7.6 An RNA Genome or a DNA Genome?

In modern life forms, genetic information is inherited by DNA replication. Genetic information embedded in a gene (a segment of a genomic DNA) is transcribed to mRNA followed by translation to the amino acid sequence. Some DNA replication-related proteins found in *Bacteria* are not evolutionarily related to those found in *Archaea* and *Eukarya*. In contrast, most of translation-related proteins are homologous among the three major domains of life. Therefore, Mushegian and Koonin suggested that *C. commonote* had an RNA genome (Mushegian and Koonin 1996). Forterre (2006) thought that DNA replication first developed in a virus and evolved in the virus world. Then, three different viruses independently infected an ancient archaeon, bacterium, and eukaryote, and therefore different DNA polymerases were integrated into genome replication mechanisms among the three domains.

Dissected tRNAs were found in *Nanoarchaeum equitans* (Randau et al. 2005), a species located near the root of the archaea tree. Di Giulio (2006) assumed that the dissected tRNAs were a primitive trait and that *C. commonote* had a fragmented genome made up of RNA. However, Martin and coworkers (Weiss et al. 2016) indicated that several genes for DNA replication-related proteins and DNA-binding proteins were included in the 355-gene set potentially present in *C. commonote*. Moreover, given the facts that all known organisms have a DNA genome and that no RNA genome organism has been found except for some RNA viruses, the DNA genome hypothesis seems to be more convincing. In addition, cyclic structures found for both modern bacterial and archaeal genomic DNAs imply that *C. commonote* also had a cyclic DNA genome (Yamagishi et al. 1998).

7.7 Cell Membrane

A cell membrane divides the inside and outside of a cell and is therefore essential for life. All modern organisms have phospholipids and hydrocarbon chains as the main components of the membrane. Most bacteria and eukaryotes have ester lipids

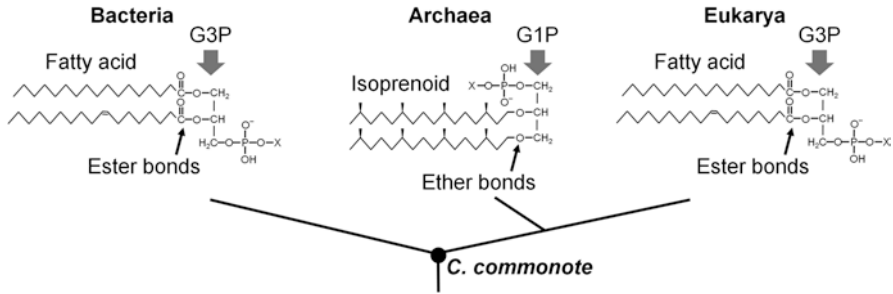


Fig. 7.4 Structures of polar lipids that are the major components of cell membrane. Generally, bacteria and eukaryotes have ester lipids with long fatty acid chains, whereas archaea have ether lipids with isoprenoid chains. The glycerol backbone in the phospholipids of *Bacteria* and *Eukarya* is G3P, whereas that of *Archaea* is G1P, the enantiomer of G3P

with long fatty acid chains, whereas archaea generally have ether lipids with isoprenoid chains. Moreover, the glycerol backbone in the phospholipids of *Bacteria* and *Eukarya* is *sn*-glycerol-3-phosphate (G3P), whereas that of *Archaea* is *sn*-glycerol-1-phosphate (G1P) (Fig. 7.4). G1P is the enantiomer of G3P. G1P and G3P are produced from dihydroxyacetone phosphate (DHAP) by G1P dehydrogenase (G1PDH) and G3P dehydrogenase (G3PDH), respectively. The two dehydrogenases are evolutionally unrelated (Pereto et al. 2004). The stereochemistry of *C. commonote*'s membrane lipid has long been unknown. Recently, Yokobori and coworkers (2016) built molecular phylogenetic trees of archaeal G1PDH and its bacterial homologues and of two subfamilies of G3PDH. They also analyzed the distribution of genes encoding those proteins among *Archaea* and *Bacteria*. Their analysis implied that the G1PDH homologue found in *Bacteria* is a subgroup of archaeal G1PDH and the bacterial common ancestor did not have any G1PDH. In contrast, it is likely that archaeal G3PDHs have a deep origin in archaeal lineages, and therefore the archaeal common ancestor possessed a G3PDH. Therefore, *C. commonote* most likely had G3PDH, and its cellular membrane would have been composed of the polar lipid with G3P. In addition, *C. commonote* may have had a fatty acid membrane lipid because some archaeal species, in addition to *Bacteria* and *Eukarya*, have genes that are homologous with those involved in fatty acid metabolism (Dibrova et al. 2014).

7.8 Reconstruction of a Protein Possessed by *C. commonote*

Currently, ancestral sequences of some proteins likely possessed by *C. commonote* or other ancestral organisms can be inferred using *in silico* methods. The key steps for the procedure are illustrated in Fig. 7.5. Extant homologous protein sequences collected from public databases are aligned to construct a multiple sequence alignment using an alignment algorithm such as MAFFT (Katoh and Standley 2013). If

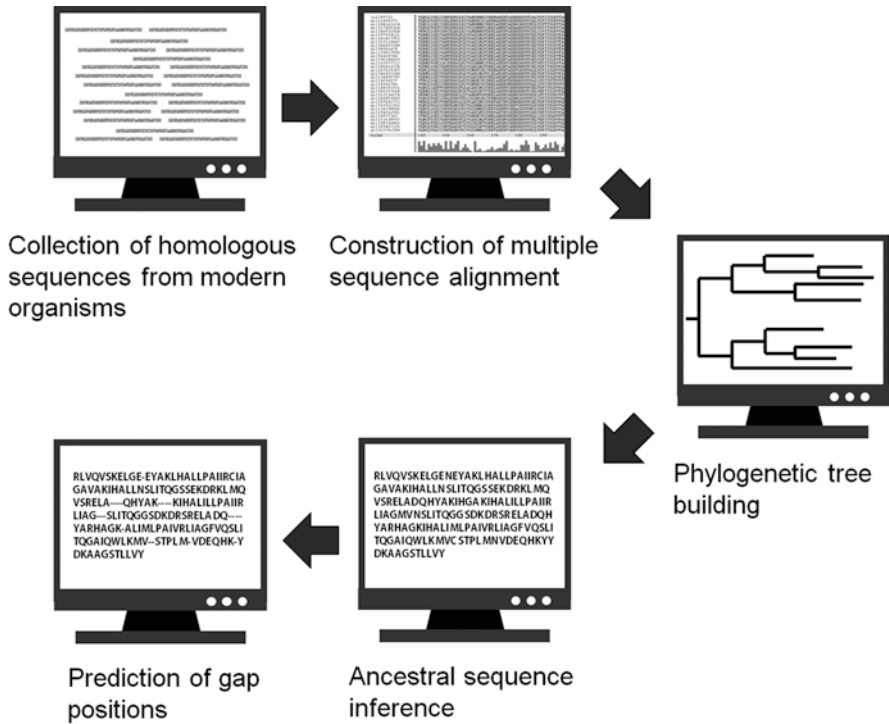


Fig. 7.5 The procedure to infer an ancestral protein sequence

necessary, insertion/gap positions are manually corrected. Then, a phylogenetic tree is built from the resulting alignment. Using the topology of the tree and the homologous sequences contained in the tree, ancestral amino acid sequences are computed based on either a homogeneous or a nonhomogeneous evolution model. The homogeneous model assumes constant global amino acid compositions in proteins throughout the tree (Yang 1997), whereas the nonhomogeneous model allows variability of global amino acid compositions at different times and lineages of the tree (Galtier et al. 1999; Blanquart and Lartillot 2008). Finally, gap positions in the ancestral sequence are predicted using a program such as GASP (Edwards and Shields 2004). Detailed procedures for the ancestral sequence reconstruction are described in several reviews by others (Thornton 2004; Gaucher et al. 2010; Merkl and Sterner 2016).

7.9 Environmental Temperature

There has been a long-running argument about the environmental temperature of *C. commonote*. In one of the most referenced species trees based on ribosomal RNA sequences of modern organisms, hyperthermophilic bacteria and archaea are located

near the root of the tree (Achenbach-Richter et al. 1988; Woese 1987), suggesting that bacterial and archaeal common ancestors were hyperthermophilic (Pace 1991; Stetter 2006). The principle of parsimony therefore suggests a hyperthermophilic *C. commonote*. Guanine plus cytosine contents in ancestral states of ribosomal RNAs and amino acid compositions of ancestral proteins were inferred based on in silico reconstruction of ancestral ribosomal RNAs and proteins. Some of those studies suggest that *C. commonote* was a mesophilic organism (Galtier et al. 1999; Boussau et al. 2008), whereas other studies support the idea that *C. commonote* was thermophilic or hyperthermophilic (Di Giulio 2000, 2003; Brooks et al. 2004). Thus, no consistent conclusion has yet been obtained from the in silico studies.

Pioneer in vitro experiments for testing the hyperthermophilicity of *C. commonote* were conducted by Yamagishi and coworkers (Miyazaki et al. 2001; Watanabe et al. 2006; Shimizu et al. 2007). They computed an ancestral amino acid sequence of a protein plausibly possessed by *C. commonote*. Then, they replaced a few amino acids in the sequence of a modern thermophilic descendant with those found in the reconstructed ancestral sequence and compared the thermal stabilities of the mutants with that of the wild-type protein. The mutants showed a tendency to be more thermostable than the wild-type thermophilic protein, suggesting that *C. commonote* was a hyperthermophilic organism.

Entire ancestral amino acid sequences have also been synthesized in vitro and then characterized. Generally, if the environmental temperature of an organism is higher, the thermal stability of a protein possessed by the organism is also greater. Indeed, the midpoint denaturation temperature of a nucleoside diphosphate kinase (NDK) correlates well with its host's environmental temperature (Akanuma et al. 2013a). The midpoint denaturation temperatures of reconstructed NDKs plausibly possessed by *C. commonote* suggest that it was a hyperthermophile that thrived at a temperature above 90 °C (Akanuma et al. 2015). Other empirical studies also support a thermophilic or hyperthermophilic ancestry of life (Gaucher et al. 2003, 2008; Butzin et al. 2013). The recently reported 355-gene set potentially possessed by *C. commonote* included a gene for reverse gyrase, a hyperthermophile-specific protein, which further supports the theory that *C. commonote* was a hyperthermophile (Weiss et al. 2016). A discussion of the hyperthermophilicity of ancient organisms is presented in greater detail elsewhere (Akanuma 2017).

7.10 Can We Trace Back Ancestry Before *C. commonote*?

Reconstruction of ancestral genes or proteins before the time of *C. commonote* could be a powerful approach to better understand ancestral life and its environment (Cornish-Bowden and Cardenas 2017). A composite tree consisting of sequences for three or more paralogous genes or proteins that may have diverged before the time of *C. commonote* is required to infer an ancestral sequence that was possessed by an ancient organism prior to *C. commonote*. Such composite trees have been built for aminoacyl-tRNA synthetases (Brown and Doolittle 1995; Fournier et al.

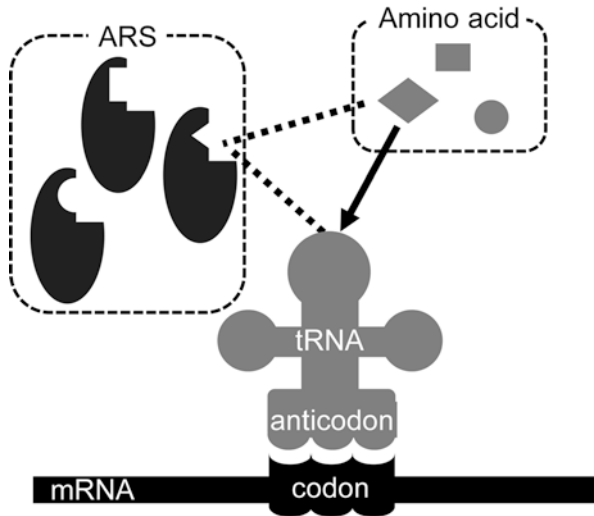


Fig. 7.6 Aminoacyl-tRNA synthetases (ARSs) attach amino acids to cognate tRNA molecules. ARSs are essential enzymes for protein biosynthesis and responsible for the accurate translation of the genetic code. Most existing organisms possess a set of 20 ARSs, which specifically recognize and charge amino acids to their cognate tRNA molecules

2011). What can we learn from reconstructed aminoacyl-tRNA synthetases? It is reasonably assumed that *C. commonote* synthesized proteins with the same 20 amino acids commonly used in modern organisms. However, it has been proposed that the earliest protein synthesis system involved a limited set of amino acids that were abundant in the primitive environment (Miller 1953; Cleaves 2010; Philip and Freeland 2011; Longo et al. 2013). Because aminoacyl-tRNA synthetase is responsible for amino acid specificity in the translation system (Fig. 7.6), reconstruction and characterization of ancestral aminoacyl-tRNA synthetases would shed light on how the common translation system involving 20 amino acids developed during the early stage of evolution. Currently, studies in my group are being conducted to address these questions.

7.11 Conclusion

C. commonote is thought to have been a unique cell or a population of cells that shared a gene set essential for living approximately 3800 Mya. The universal ancestor may have had a cyclic DNA genome and basic metabolic pathways. The possible gene content of *C. commonote* suggests that it was an anaerobic autotroph depending on H_2 and CO_2 . The cell may have been surrounded by a stable cell membrane similar to those found in modern bacteria and eukaryotes. Ancestral protein reconstruction studies suggest that the universal ancestor was a hyperthermophile that

thrived at a very high temperature. However, these conclusions will need to be revised when new microbes with ancient traits are discovered or when more genes and proteins present in *C. commonote* are reconstructed and characterized.

References

- Achenbach-Richter L, Gupta R, Zillig W, Woese CR (1988) Rooting the archaeobacterial tree: the pivotal role of *Thermococcus celer* in archaeobacterial evolution. *Syst Appl Microbiol* 10:231–240
- Akanuma S (2017) Characterization of reconstructed ancestral proteins suggests a change in temperature of the ancient biosphere. *Life (Basel, Switzerland)* 7(3). doi:<https://doi.org/10.3390/life7030033>
- Akanuma S, Nakajima Y, Yokobori S, Kimura M, Nemoto N, Mase T, Miyazono K, Tanokura M, Yamagishi A (2013a) Experimental evidence for the thermophilicity of ancestral life. *Proc Natl Acad Sci USA* 110(27):11067–11072. <https://doi.org/10.1073/pnas.1308215110>
- Akanuma S, Yokobori S, Yamagishi A (2013b) Comparative genomics of thermophilic bacteria and archaea. In: Satyanarayana T, Litterchild J, Kawarabayasi Y (eds) *Thermophilic microbes in environmental and industrial biotechnology*. Springer, Dordrecht, pp 331–349
- Akanuma S, Yokobori S, Nakajima Y, Bessho M, Yamagishi A (2015) Robustness of predictions of extremely thermally stable proteins in ancient organisms. *Evol Intl J Org Evol* 69(11):2954–2962. <https://doi.org/10.1111/evo.12779>
- 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(47):14518–14521. <https://doi.org/10.1073/pnas.1517557112>
- Blanquart S, Lartillot N (2008) A site- and time-heterogeneous model of amino acid replacement. *Mol Biol Evol* 25(5):842–858. <https://doi.org/10.1093/molbev/msn018>
- Boussau B, Blanquart S, Necsulea A, Lartillot N, Gouy M (2008) Parallel adaptations to high temperatures in the Archaean eon. *Nature* 456(7224):942–945. <https://doi.org/10.1038/nature07393>
- Brooks DJ, Fresco JR, Singh M (2004) A novel method for estimating ancestral amino acid composition and its application to proteins of the Last Universal Ancestor. *Bioinformatics (Oxford, England)* 20(14):2251–2257. <https://doi.org/10.1093/bioinformatics/bth235>
- Brown JR, Doolittle WF (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proc Natl Acad Sci USA* 92(7):2441–2445
- Butzin NC, Lapierre P, Green AG, Swithers KS, Gogarten JP, Noll KM (2013) Reconstructed ancestral Myo-inositol-3-phosphate synthases indicate that ancestors of the Thermococcales and Thermotoga species were more thermophilic than their descendants. *PLoS One* 8(12):e84300. <https://doi.org/10.1371/journal.pone.0084300>
- Cavalier-Smith T (2002) The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int J Syst Evol Microbiol* 52(Pt 1):7–76. <https://doi.org/10.1099/00207713-52-1-7>
- Cavalier-Smith T (2006a) Cell evolution and Earth history: stasis and revolution. *Philos Trans R Soc Lond Ser B Biol Sci* 361(1470):969–1006. <https://doi.org/10.1098/rstb.2006.1842>
- Cavalier-Smith T (2006b) Rooting the tree of life by transition analyses. *Biol Direct* 1:19. <https://doi.org/10.1186/1745-6150-1-19>
- Cavalier-Smith T (2010) Deep phylogeny, ancestral groups and the four ages of life. *Philos Trans R Soc Lond Ser B Biol Sci* 365(1537):111–132. <https://doi.org/10.1098/rstb.2009.0161>
- Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science (New York, NY)* 311(5765):1283–1287. <https://doi.org/10.1126/science.1123061>

- Cleaves HJ 2nd (2010) The origin of the biologically coded amino acids. *J Theor Biol* 263(4):490–498. <https://doi.org/10.1016/j.jtbi.2009.12.014>
- Cornish-Bowden A, Cardenas ML (2017) Life before LUCA. *J Theor Biol.* <https://doi.org/10.1016/j.jtbi.2017.05.023>
- Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM (2008) The archaeobacterial origin of eukaryotes. *Proc Natl Acad Sci USA* 105(51):20356–20361. <https://doi.org/10.1073/pnas.0810647105>
- Darwin C (1859) *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.* Murray, London
- Di Giulio M (2000) The universal ancestor lived in a thermophilic or hyperthermophilic environment. *J Theor Biol* 203(3):203–213. <https://doi.org/10.1006/jtbi.2000.1086>
- Di Giulio M (2003) The universal ancestor and the ancestor of bacteria were hyperthermophiles. *J Mol Evol* 57(6):721–730. <https://doi.org/10.1007/s00239-003-2522-6>
- Di Giulio M (2006) The non-monophyletic origin of the tRNA molecule and the origin of genes only after the evolutionary stage of the last universal common ancestor (LUCA). *J Theor Biol* 240(3):343–352. <https://doi.org/10.1016/j.jtbi.2005.09.023>
- Dibrova DV, Galperin MY, Mulikidjanian AY (2014) Phylogenomic reconstruction of archaeal fatty acid metabolism. *Environ Microbiol* 16(4):907–918. <https://doi.org/10.1111/1462-2920.12359>
- Dodd MS, Papineau D, Grenne T, Slack JF, Rittner M, Pirajno F, O’Neil J, Little CT (2017) Evidence for early life in Earth’s oldest hydrothermal vent precipitates. *Nature* 543(7643):60–64. <https://doi.org/10.1038/nature21377>
- Doolittle WF (1999) Phylogenetic classification and the universal tree. *Science (New York, NY)* 284(5423):2124–2129
- Doolittle WF, Brown JR (1994) Tempo, mode, the progenote, and the universal root. *Proc Natl Acad Sci USA* 91(15):6721–6728
- Edwards RJ, Shields DC (2004) GASP: gapped ancestral sequence prediction for proteins. *BMC Bioinforma* 5:123. <https://doi.org/10.1186/1471-2105-5-123>
- Feng DF, Cho G, Doolittle RF (1997) Determining divergence times with a protein clock: update and reevaluation. *Proc Natl Acad Sci USA* 94(24):13028–13033
- 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 USA* 103(10):3669–3674. <https://doi.org/10.1073/pnas.0510333103>
- Fournier GP, Andam CP, Alm EJ, Gogarten JP (2011) Molecular evolution of aminoacyl tRNA synthetase proteins in the early history of life. *Orig Life Evol Biosph J Int Soc Study Orig Life* 41(6):621–632. <https://doi.org/10.1007/s11084-011-9261-2>
- Fournier GP, Andam CP, Gogarten JP (2015) Ancient horizontal gene transfer and the last common ancestors. *BMC Evol Biol* 15:70. <https://doi.org/10.1186/s12862-015-0350-0>
- Furukawa R, Nakagawa M, Kuroyanagi T, Yokobori SI, Yamagishi A (2017) Quest for ancestors of eukaryal cells based on phylogenetic analyses of aminoacyl-tRNA synthetases. *J Mol Evol* 84(1):51–66. <https://doi.org/10.1007/s00239-016-9768-2>
- Galtier N, Tourasse N, Gouy M (1999) A nonhyperthermophilic common ancestor to extant life forms. *Science (New York, NY)* 283(5399):220–221
- Gaucher EA, Thomson JM, Burgan MF, Benner SA (2003) Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* 425(6955):285–288. <https://doi.org/10.1038/nature01977>
- Gaucher EA, Govindarajan S, Ganesh OK (2008) Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* 451(7179):704–707. <https://doi.org/10.1038/nature06510>
- Gaucher EA, Kratzer JT, Randall RN (2010) Deep phylogeny – how a tree can help characterize early life on Earth. *Cold Spring Harb Perspect Biol* 2(1):a002238. <https://doi.org/10.1101/cshperspect.a002238>
- Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF, Poole RJ, Date T, Oshima T et al (1989) Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proc Natl Acad Sci USA* 86(17):6661–6665

- Grossin M, Hobbs JK, Szollosi GJ, Gribaldo S, Arcus VL, Gouy M (2015) Toward more accurate ancestral protein genotype-phenotype reconstructions with the use of species tree-aware gene trees. *Mol Biol Evol* 32(1):13–22. <https://doi.org/10.1093/molbev/msu305>
- Harris JK, Kelley ST, Spiegelman GB, Pace NR (2003) The genetic core of the universal ancestor. *Genome Res* 13(3):407–412. <https://doi.org/10.1101/gr.652803>
- Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T (1989) Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc Natl Acad Sci USA* 86(23):9355–9359
- Kandler O (1995) Cell wall biochemistry in Archaea and its phylogenetic implications. *J Biol Phys* 20(1):165–169. <https://doi.org/10.1007/bf00700433>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772–780. <https://doi.org/10.1093/molbev/mst010>
- Lake JA, Servin JA, Herbold CW, Skophammer RG (2008) Evidence for a new root of the tree of life. *Syst Biol* 57(6):835–843. <https://doi.org/10.1080/10635150802555933>
- Lake JA, Skophammer RG, Herbold CW, Servin JA (2009) Genome beginnings: rooting the tree of life. *Philos Trans R Soc Lond Ser B Biol Sci* 364(1527):2177–2185. <https://doi.org/10.1098/rstb.2009.0035>
- Longo LM, Lee J, Blaber M (2013) Simplified protein design biased for prebiotic amino acids yields a foldable, halophilic protein. *Proc Natl Acad Sci USA* 110(6):2135–2139. <https://doi.org/10.1073/pnas.1219530110>
- Merkel R, Sterner R (2016) Ancestral protein reconstruction: techniques and applications. *Biol Chem* 397(1):1–21. <https://doi.org/10.1515/hsz-2015-0158>
- Miller SL (1953) A production of amino acids under possible primitive earth conditions. *Science (New York, NY)* 117(3046):528–529
- Miyazaki J, Nakaya S, Suzuki T, Tamakoshi M, Oshima T, Yamagishi A (2001) Ancestral residues stabilizing 3-isopropylmalate dehydrogenase of an extreme thermophile: experimental evidence supporting the thermophilic common ancestor hypothesis. *J Biochem* 129(5):777–782
- Mushegian AR, Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci USA* 93(19):10268–10273
- Nisbet EG, Sleep NH (2001) The habitat and nature of early life. *Nature* 409(6823):1083–1091. <https://doi.org/10.1038/35059210>
- Nutman AP, Bennett VC, Friend CR, Van Kranendonk MJ, Chivas AR (2016) Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. *Nature* 537(7621):535–538. <https://doi.org/10.1038/nature19355>
- Pace NR (1991) Origin of life – facing up to the physical setting. *Cell* 65(4):531–533
- Pereto J, Lopez-García P, Moreira D (2004) Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem Sci* 29(9):469–477. <https://doi.org/10.1016/j.tibs.2004.07.002>
- Philip GK, Freeland SJ (2011) Did evolution select a nonrandom “alphabet” of amino acids? *Astrobiology* 11(3):235–240. <https://doi.org/10.1089/ast.2010.0567>
- Randau L, Munch R, Hohn MJ, Jahn D, Soll D (2005) Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5'- and 3'-halves. *Nature* 433(7025):537–541. <https://doi.org/10.1038/nature03233>
- Raymann K, Brochier-Armanet C, Gribaldo S (2015) The two-domain tree of life is linked to a new root for the Archaea. *Proc Natl Acad Sci USA* 112(21):6670–6675. <https://doi.org/10.1073/pnas.1420858112>
- Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, Darling A, Malfatti S, Swan BK, Gies EA, Dodsorth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu WT, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499(7459):431–437. <https://doi.org/10.1038/nature12352>
- Rivera MC, Lake JA (1992) Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science (New York, NY)* 257(5066):74–76

- Rosing MT (1999) ^{13}C -depleted carbon microparticles in >3700-Ma sea-floor sedimentary rocks from West Greenland. *Science* (New York, NY) 283(5402):674–676
- Schopf JW (1993) Microfossils of the Early Archean Apex chert: new evidence of the antiquity of life. *Science* (New York, NY) 260:640–646
- Shen Y, Buick R, Canfield DE (2001) Isotopic evidence for microbial sulphate reduction in the early Archean era. *Nature* 410(6824):77–81. <https://doi.org/10.1038/35065071>
- Shimizu H, Yokobori S, Ohkuri T, Yokogawa T, Nishikawa K, Yamagishi A (2007) Extremely thermophilic translation system in the common ancestor commonote: ancestral mutants of Glycyl-tRNA synthetase from the extreme thermophile *Thermus thermophilus*. *J Mol Biol* 369(4):1060–1069. <https://doi.org/10.1016/j.jmb.2007.04.001>
- Stetter KO (2006) Hyperthermophiles in the history of life. *Philos Trans R Soc Lond Ser B Biol Sci* 361(1474):1837–1842.; discussion 1842–1833. <https://doi.org/10.1098/rstb.2006.1907>
- Theobald DL (2010) A formal test of the theory of universal common ancestry. *Nature* 465(7295):219–222. <https://doi.org/10.1038/nature09014>
- Thornton JW (2004) Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat Rev Genet* 5(5):366–375. <https://doi.org/10.1038/nrg1324>
- Ueno Y, Yamada K, Yoshida N, Maruyama S, Isozaki Y (2006) Evidence from fluid inclusions for microbial methanogenesis in the early Archean era. *Nature* 440(7083):516–519. <https://doi.org/10.1038/nature04584>
- Wacey D, McLoughlin N, Green OR, Parnell J, Stoakes CA, Brasier MD (2006) The ~3.4 billion-year-old Strelley Pool Sandstone: a new window into early life on Earth. *Int J Astrobiol* 5(04):333. <https://doi.org/10.1017/s1473550406003466>
- Watanabe K, Ohkuri T, Yokobori S, Yamagishi A (2006) Designing thermostable proteins: ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree. *J Mol Biol* 355(4):664–674. <https://doi.org/10.1016/j.jmb.2005.10.011>
- Weiss MC, Sousa FL, Mmjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF (2016) The physiology and habitat of the last universal common ancestor. *Nat Microbiol* 1(9):16116. <https://doi.org/10.1038/nmicrobiol.2016.116>
- Williams TA, Foster PG, Cox CJ, Embley TM (2013) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* 504(7479):231–236. <https://doi.org/10.1038/nature12779>
- Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51(2):221–271
- Woese CR, Fox GE (1977) The concept of cellular evolution. *J Mol Evol* 10(1):1–6
- Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 87(12):4576–4579
- Yamagishi A, Kon T, Takahashi G, Oshima T (1998) From the common ancestor of all living organisms to protoeukaryotic cell. In: Wiegel J, Adams MWW (eds) *Thermophiles: the keys to molecular evolution and the origin of life?* Taylor & Francis, London
- Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci CABIOS* 13(5):555–556
- Yokobori SI, Nakajima Y, Akanuma S, Yamagishi A (2016) Birth of archaeal cells: molecular phylogenetic analyses of G1P dehydrogenase, G3P dehydrogenases, and glycerol kinase suggest derived features of archaeal membranes having G1P polar lipids. *Archaea* (Vancouver, BC) 2016:1802675. <https://doi.org/10.1155/2016/1802675>
- Yutin N, Makarova KS, Mekhedov SL, Wolf YI, Koonin EV (2008) The deep archaeal roots of eukaryotes. *Mol Biol Evol* 25(8):1619–1630. <https://doi.org/10.1093/molbev/msn108>

Chapter 8

Eukaryotes Appearing



Shin-ichi Yokobori and Ryutaro Furukawa

Abstract The appearance of eukaryotic cells was a major step in the evolution of terrestrial life. Recent phylogenetic analyses indicate that the Eukaryotes appeared from the Archaeobacteria rather than being a distinct domain from Archaeobacteria and Eubacteria. The Asgard archaeal group, which shares genes that are otherwise unique to Eukaryotes, has been suggested to be the closest relative to Eukaryotes. However, eukaryotic genes have also been shown to have originated from diverse groups in the Archaeobacteria and Eubacteria. Asgard archaeon-like Archaea (Archaeobacteria) may have been the host for endosymbiosis of the mitochondrial ancestor (Alphaproteobacteria) and might have been the ancestor of Eukaryotes; nevertheless, horizontal gene transfer from various lineages of Archaeobacteria and Eubacteria also appear to have played an important role in the evolution of Eukaryotes.

Keywords Asgard group · Mitochondria · Horizontal gene transfer

8.1 Introduction

Cellular organisms are classified into two groups, namely, prokaryotes and eukaryotes. Eukaryotes have a nucleus containing chromosomes, a complex membrane system including endoplasmic reticulum (ER) and Golgi apparatus, and double-membrane organelles, such as mitochondria and plastids. In contrast, prokaryotes have no apparent structures in which chromosome(s) are stored. There are two distinct groups of prokaryotes, namely, Bacteria (Eubacteria) and Archaea (Archaeobacteria) (Woese 1987; Woese et al. 1990). In general, prokaryotic cells are much smaller and simpler than eukaryotic cells: eukaryotic cells are 10 to several 100 times larger than typical prokaryotic cells.

S.-i. Yokobori (✉) · R. Furukawa
Laboratory of Bioengineering, Department of Applied Life Sciences, School of Life Sciences,
Tokyo University of Pharmacy and Life Sciences, Tokyo, Japan
e-mail: yokobori@toyaku.ac.jp

Paleobiological records on early eukaryotes are not preserved well and are hard to interpret; nevertheless, they suggest that eukaryotes appeared ~1.7 billion years ago (Dacks et al. 2016). Eme et al. (2014) dated the age of the eukaryotic common ancestor as 1000–1900 million years ago based on molecular phylogenetic analyses. These dates are much later than the suggested date of appearance of Archaea (Dacks et al. 2016) based on chemical fossil records, such as isotopically light methane (Ueno et al. 2008).

The serial endosymbiotic hypothesis for the origin of eukaryotes proposed by Margulis is widely accepted in the scientific community (Sagan 1967; Margulis 1970). In this hypothesis, an endosymbiotic origin of mitochondria (alphaproteobacterial origin), plastids (cyanobacterial origin), and flagella has been proposed. Molecular phylogenetic analyses have supported the endosymbiotic origin of mitochondria and plastids (Anderson et al. 1998; Nelissen et al. 1995). Pisani et al. (2007) reported that eukaryotic genes show high similarities with their counterparts in the Alphaproteobacteria, *Cyanobacteria*, and *Thermoplasmata*; they used a supertree method, which will be explained later in this chapter, to identify relationships. The high similarity with Alphaproteobacteria and *Cyanobacteria* can be explained by the acquisition of mitochondria and plastids via endosymbiosis. Although amitochondrial eukaryotes (eukaryotes lacking mitochondria) have been found (Roger et al. 1998; Karnkowska et al. 2016), phylogenetic and genome analyses suggest that the absence of mitochondria is secondary and that they are descendants of eukaryotes with mitochondria (Karnkowska et al. 2016). Therefore, having mitochondria is a symplesiomorphic (= shared ancestral) character of extant eukaryotes.

The acquisition of mitochondria appears to have been one of the important events in the formation of eukaryotes, and identification of the host organisms of the mitochondrial ancestor is essential to clarify how eukaryotes emerged. It is also important to resolve the genesis of subcellular structures (organelles) other than mitochondria and plastids (e.g., López-García and Moreira 2015). With a few exceptions, such as *Planctomycetes* (Fuerst and Sagulenko 2011), subcellular structures with membranes are not developed in prokaryotic cells.

Progress in genomic and metagenomic methods over the last two decades has enabled the comparison of genes and genomes from a wide range of organisms (Ciccarelli et al. 2006; Castelle et al. 2015). In particular, metagenomic studies have allowed the investigation of unculturable eubacteria and archaeobacteria (e.g., Elkin et al. 2008; Nunoura et al. 2011; Rinke et al. 2013; Spang et al. 2015; Castelle et al. 2015; Hug et al. 2016; Zaremba-Niedzwiedzka et al. 2017). Genome-based studies have revealed new aspects of the origin of Eukaryotes (eukaryogenesis) (e.g., Guy and Ettema 2011; Spang et al. 2015; Hug et al. 2016; Zaremba-Niedzwiedzka et al. 2017).

In this chapter, we discuss the origin of the eukaryotic cell using information from recent phylogenomic studies.

8.2 “Tree of Life” and Eukaryogenesis

In the late 1980s, it was established that prokaryotes consist of two phylogenetically and genetically distinct groups, the Eubacteria and the Archaeobacteria (Woese 1987). As a consequence, cellular life was reclassified into three domains, Bacteria (or Eubacteria), Archaea (or Archaeobacteria), and Eucarya (or Eukaryota Eukaryotes) (Woese et al. 1990). In this chapter we will use the terms Eubacteria, Archaeobacteria, and Eukaryotes. Some of the characteristics of Eukaryotes, Archaeobacteria, and Eubacteria are listed in Table 8.1. As can be seen, Eukaryotes share some characteristics with Archaeobacteria and other characteristics with Eubacteria. This suggests that Eukaryotes have acquired both archaeobacterial and eubacterial characteristics (genes), as discussed later.

If the Eubacteria, Archaeobacteria, and Eukaryotes are monophyletic groups, then the relationship among the three groups will take the form of an unrooted tree (Fig. 8.1a). When a root is placed on the eubacterial branch (asterisk 1), then Archaeobacteria and Eukaryotes are sister groups (Fig. 8.1b). This rooting had been proposed by Iwabe et al. (1989) from a phylogenetic analysis of composite data on translational elongation factor (EF) Tu/1 α and EF G/2. Woese et al. (1990) placed the position of the root on the small subunit ribosomal RNA sequences (ssu rRNA or 16S/18S rRNA) “Tree of Life,” based on the root position suggested by Iwabe et al. (1989). Harris et al. (2003) performed phylogenetic analyses of 80 protein gene clusters conserved among the three domains and suggested that trees based on 50 of the 80 protein gene clusters showed similar topologies to the ssu rRNA gene tree (three domain hypothesis). Yutin et al. (2008) analyzed Archaeobacteria-originating protein genes in Eukaryotes and found that most appeared as sister groups of archaeobacterial genes. These studies also suggested that the Archaeobacteria and Eukaryotes are distinct groups. In this view, Eukaryotes share a common ancestor with Archaeobacteria, although these two groups evolved independently after their separation.

There are other possible locations for the root of a three-domain tree. For example, the root may be on the archaeobacterial branch with a close relationship between Eubacteria and Eukaryotes (Fig. 8.1c). This hypothesis cannot be ignored since various eukaryotic genes are known to be of eubacterial origin rather than archaeobacterial origin (e.g., Thierygart et al. 2012; Ku et al. 2015). Another possibility is that the root is on the eukaryotic branch, suggesting deep diversity between eukaryotes and prokaryotes (Fig. 8.1d). Hypotheses to explain this potential relationship have been put forward (e.g., Harish and Kurland 2017); however, we will not consider these further in this chapter, since the date of appearance of Eukaryotes is thought to have been much later than that of Eubacteria and Archaeobacteria (Dacks et al. 2016).

Table 8.1 Comparison of characteristics among Eukaryotes, Archaeobacteria (Archaea), and Eubacteria (Bacteria)

Characteristics	Eukaryotes (nucleus/cytoplasm)	Archaeobacteria	Eubacteria
Size of cell	Several to 100 μm	0.5–10 μm	0.5–10 μm
Nucleus separated by membrane from cytoplasm	Yes	No	No ^a
Lipids in cellular membrane	Ester bond (<i>sn</i> -1,2 position), straight chain hydrocarbon (fatty acid)	Ether bond (<i>sn</i> -2,3 position), branched chain hydrocarbon (isoprenoid)	Ester bond (<i>sn</i> -1,2 position), straight chain hydrocarbon (fatty acid)
Endocytosis (phagocytosis and pinocytosis)	Present	Absent	Absent ^a
Actin and its relatives	Present	Present (limited) (MreB)	Present (limited) (MrsB)
Tubulin and its relatives	Present	Present (limited) (FtsZ)	Present (limited) (FtsZ)
Intermediate filament	Present	Present (limited)	Present (limited)
Organelles	Present	Absent	Absent
Replication, genome and gene structure	Structure of chromosomes	Linear	Circular (with exception)
	DNA polymerase in replication	Family B	Family B (+D in some species)
	DNA binding protein	Histone	Archaeobacterial histone ^b
	Operon	Absent	Present
	Introns in mRNA	Present	Absent
Transcription	Cap at the 5' end of mRNA	Present	Absent
	Poly A at the 3' end of mRNA	Present	Present (degradative)
	RNA polymerases	3 species (Pol I, Pol II, Pol III)	1 species (Pol II type)
	Initiation system of transcription	Preinitiation complex of transcription (TFIIIB, TBP)	TFB-TBP
	Initiation tRNA	Methionine	Methionine
Translation	Introns in tRNA	Present	Present, but few
	Size of ribosome	80S (60S + 40S)	70S (50S + 30S)
Methanogenesis	Absent	Present	Absent
Photosynthesis	Absent (but present in plastid (chloroplast))	Absent	Present

Modified from Borton et al. (2007); some characteristics have been added

^aThere are exceptions (*Planctomyces*)

^bThere are archaeobacterial species without archaeobacterial histone

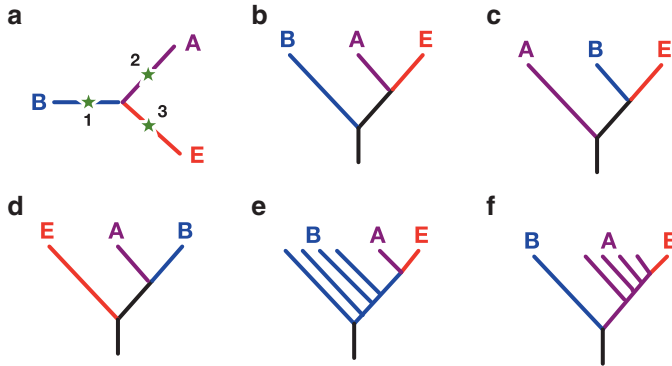


Fig. 8.1 Possible relationship among Eukaryotes (Eucarya), Archaeobacteria (Archaea), and Eubacteria (Bacteria). “A” stands for Archaeobacteria. “B” stands for Eubacteria. “E” stands for Eukaryotes. (a) Unrooted tree showing the relationship among three monophyletic domains. Asterisks show the possible positions of the root (the position of Commonote). (b–d) Three-domain model with different positions of root (position of Commonote) (Rooted trees). (b) The root is on the bacterial branch. Archaeobacteria and Eukaryotes form a group (Woese et al. 1990). (c) The root is on the archaeobacterial branch. Eubacteria and Eukaryotes form a group. (d) The root is on the eukaryotic branch. Prokaryotes form a group. (e.) Root within a eubacterial group. Archaeobacteria and Eukaryotes appear with eubacterial origins (Cavalier-Smith 2002) (Neomura hypothesis). (f) Two-domain model similar to “Eocyte hypothesis” (Lake et al. 1984; Rivera and Lake 1992). The root is placed between Eubacteria and Archaeobacteria. Eukaryotes from a group with an archaeobacterial lineage; therefore Archaeobacteria is a paraphyletic group if Eukaryotes are excluded

Each of the domains is not necessarily monophyletic. The tree presented in Fig. 8.1e is an example of such a case (paraphyly in Eubacteria). In this tree, the last universal common ancestor was a eubacterial group, and both Archaeobacteria and Eukaryotes share a common ancestor. This idea has been discussed in detail in the neomura hypothesis of Cavalier-Smith (2002). Briefly, in this hypothesis, the root of life is thought to be within Gram-negative bacteria; after the appearance of Gram-positive bacteria, a common ancestor of Archaeobacteria and Eukaryotes is thought to have originated from a lineage of *Actinobacteria* within the Gram-positive bacterial group. Archaeobacteria and Eukaryotes are treated as distinct monophyletic groups in this hypothesis. This hypothesis can be thought to be a variant of three-domain hypothesis because of sister group relationship of Eukaryotes to Archaeobacteria.

The tree presented in Fig. 8.1f shows a proposed paraphyletic evolution of the Archaeobacteria that accommodates the archaeobacterial origin of Eukaryotes. Lake and his colleagues proposed the “Eocyte” hypothesis based on the structure of ribosomes (Lake et al. 1984). The Eocyte is the phylum Crenarchaeota in Woese’s framework of archaeal taxonomy. Rivera and Lake (1992) analyzed EF-Tu/1 α and EF-G/2 using an indel analysis: they compared the insertion/deletion appeared in the groups of organisms to determine phylogeny. Based on this analysis, Rivera and Lake (1992) concluded that Eukaryotes are close relatives of Eocytes (Crenarchaeota), a subgroup of Archaeobacteria. In summary, several tree topologies have been proposed depending on the different method and/or data set.

Table 8.2 Selected studies on the origin of Eukaryotes based on the phylogenetic analyses of concatenated gene sequences

Publication	Number of genes	Supported model and ancestor of Eukaryotes
Ciccarelli et al. (2006)	31	3D
Cox et al. (2008)	45	2D (Crenarchaeota)
Foster et al. (2009)	41	2D (Crenarchaeota and Thaumarchaeota)
Kelly et al. (2011)	320	2D (Thaumarchaeota)
Guy and Ettema (2011) ^a	26	2D (TACK superphylum)
Williams et al. (2012)	29	2D (TACK superphylum)
Rinke et al. (2013)	38	3D
Williams and Embley (2014)	29	2D (TACK superphylum)
Spang et al. (2015) ^b	36	2D (Lokiarchaeota)
Hug et al. (2016)	16	2D (Lokiarchaeota, Thorarchaeota)
Zaremba-Niedzwiedzka et al. (2017) ^c	55	2D (Asgard superphylum)
Da Cunha et al. (2017)	36	3D

2D and 3D represent two and three domain models, respectively, with the proposed ancestor(s) of Eukaryotes in the parenthesis

^aTACK superphylum was proposed in this paper

^bDiscovery of Lokiarchaeota was reported

^cAsgard group (other than Lokiarchaeota) was discovered and proposed

The increased availability of genome sequences has allowed many more phylogenetic studies (Table 8.2). These molecular phylogenetic analyses are based on the use of two different methods. One is the “concatenated sequence” method: in this method, multiple genes are concatenated and then used for molecular phylogenetic analysis in a similar manner as a single gene. The other is the “supertree” method: in this method, multiple individual gene trees are reconstructed and then summarized to obtain a species tree. These analyses have generally supported the two-domain hypothesis (Table 8.2) where Eukaryotes are an in-group of Archaeobacteria, though the relationship between Archaeobacteria and Eukaryotes has differed among analyses using different data sets and different tree reconstruction methods (Tables 8.3). In conclusion, Eukaryotes are thought to have an archaeobacterial origin (Fig. 8.2f), although genes of eubacterial origin have also been found in Eukaryotes as well as those of archaeobacterial origin (Table 8.3). The last possibility will be discussed at the end of this chapter.

8.3 Asgard Group/TACK Superphylum as the Prokaryotic Ancestor of Eukaryotic Cells

Although Eukaryotes originated from the Archaeobacteria, there is considerable debate on which archaeal group is the closest relative of Eukaryotes (Table 8.2). Candidates are Crenarchaeota (Cox et al. 2008), a clade comprising Crenarchaeota

Table 8.3 Selected studies on the origin of Eukaryotes based on the summary of phylogenetic analyses of individual gene sequences

Publication	Distribution of ancestral eukaryotic gene
Pisani et al. (2007)	Supertree supporting Thermoplasmatales ancestry ^a
Saruhashi et al. (2008)	Bac, 82; Arc, 73 (Cren, 12; Eury, 11)
Yutin et al. (2008)	Bac, 436; Arc, 355 (Cren, 39; Eury, 52)
Thiergart et al. (2012)	Bac, 218; Arc, 168 (TACK, 84; Eury, 77; Nano, 7)
Rochette et al. (2014)	Bac, 243; Arc, 121 (TACK, 34; Eury, 70)
Ku et al. (2015)	Bac, 940; Arc, 314
Pittis and Gabaldon (2016)	Bac, 766; Arc, 196 (TACK, 71; Eury, 79)
Furukawa et al. (2017)	Bac, 7; Arc, 11 (Asgard, 1; Eury, 2; DPANN, 6)

^aThey also suggest gene flow from endosymbiont-origin organelles (mitochondria with alphaproteobacterial origin and plastids with cyanobacterial origin). *Bac* Eubacteria, *Arc* Archaeobacteria, *Cren* Crenarchaeota, *Eury* Euryarchaeota, *TACK* TACK superphylum, *Nano* Nanoarchaeota, *Asgard* Asgard group, and *DPANN* DPANN superphylum

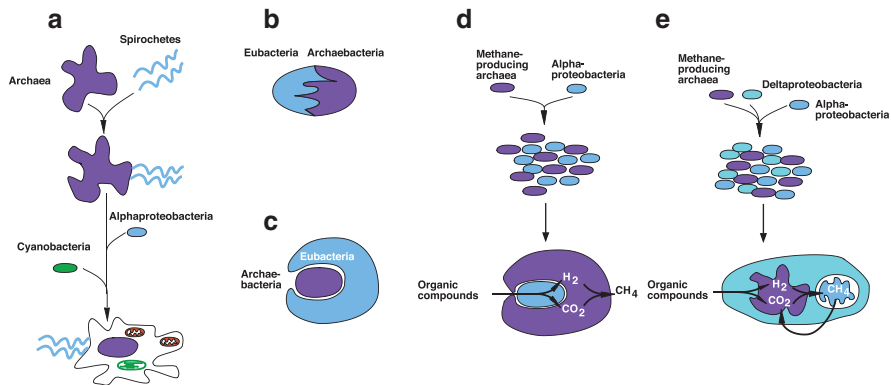


Fig. 8.2 Hypotheses on the origin of eukaryotic cells. (a) Sequential endosymbiosis hypothesis proposed by Margulis. (b) Eukaryotic cells are thought to have originated via fusion of eubacterial and archaeobacterial cells. (c) Archaeobacterial cells as the endosymbiont of eubacterial cells. Archaeobacteria become the nucleus (Lake and Rivera 1994). (d) “Hydrogen hypothesis” (Martin and Müller 1998). Eukaryotic cells are thought to have appeared through symbiosis between methane-producing Archaeobacteria and Alphaproteobacteria as the ancestry of mitochondrion and hydrogenosome. In this hypothesis, cytoplasm is thought to originate from Archaeobacteria. (e) “Syntrophy hypothesis” (López-García and Moreira 1999). This hypothesis looks similar to the “hydrogen hypothesis” but differs due to the suggested contribution by *Deltaproteobacteria* to eukaryogenesis in addition to methane-producing Archaeobacteria and Alphaproteobacteria. In this hypothesis, the cytoplasm is thought to have originated from *Deltaproteobacteria* and the nucleus from the Archaeobacteria. (Modified from Fig. 12.4 of Yokobori and Yamagishi 2013)

and Thaumarchaeota (Foster et al. 2009), within or as a sister to Thaumarchaeota (Kelly et al. 2011), Euryarchaeota (Pisani et al. 2007), and an unidentified deep archaeobacterial branch (Yutin et al. 2008; Saruhashi et al. 2008).

Guy and Ettema (2011) first proposed the TACK superphylum, composed of Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota, as the archae-

bacterial ancestor of eukaryotic cells. They pointed out that the analyzed genomes from the TACK superphylum include core Eukaryote genes such as those for cell division/vesicle formation/cell-shape determination, DNA packaging/replication/repair, ubiquitin system for protein recycling, and Eukaryote-like transcription and translation.

Subsequently, Lokiarchaeota, as a member of the TACK superphylum, was suggested as the closest archaeobacterial group to Eukaryotes (Spang et al. 2015) based on metagenome analysis and molecular phylogenetic analyses using 36 concatenated protein gene sequences. These analyses showed Lokiarchaeota had more core Eukaryote genes than other phyla in the TACK superphylum. The Lokiarchaea are suggested to be hydrogen-dependent autotrophic organisms (Sousa et al. 2016), which are consistent with them being close relatives of Eukaryotes under the hydrogen or syntrophy hypotheses (Martin and Müller 1998; López-García and Moreira 1999).

More recently, a group of Archaeobacteria was discovered, which are closely related to Lokiarchaeota and cannot be cultured; they have been named the “Asgard group” (Zaremba-Niedzwiedzka et al. 2017). The Asgard group consists of Thorarchaeota, Odinararchaeota, Heimdallarchaeota, and Lokiarchaeota. This group has been suggested to be a sister group of the TACK superphylum, and it was further suggested that Eukaryotes may have arisen from the Asgard group. The Asgard group carries various core Eukaryote genes in their genomes (Table 8.4; Spang et al. 2017; Eme et al. 2017), supporting that the Asgard groups are close relatives of Eukaryotes.

The host of mitochondria may have needed the ability to undertake endocytosis (e.g., Martin et al. 2015). Through phagocytosis, a type of endocytosis, eukaryotic cells take up large materials, such as bacterial cells, which are confined to vesicles with lipid membranes, and then lysed. Phagocytosis may be related to endosymbiosis (e.g., Yutin et al. 2009); interestingly, endosymbionts are often doubly surrounded by lipid membranes (e.g., Serbus et al. 2008): which are called outer and inner membranes, respectively. The outer membrane of double-membrane organelles, such as mitochondria and plastids, is thought to have originated from the lipid membrane generated during phagocytosis in the host (e.g., Martin et al. 2015). The Asgard group had been reported to have some genes on the remodeling of membrane (Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017. See also Table 8.4). The various proteins involved in phagocytosis and/or remodeling of membranes have been discussed elsewhere (Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017). Possible ability of membrane-remolding of Asgard archaeobacteria supports the close relationship between Asgard archaeobacteria and Eukaryotes.

Pitis and Gabaldon (2016) also suggested that eukaryotic genes with lokiarchaeal affinity seem to have participated in later phases of eukaryogenesis among those with archaeobacterial affinity. This suggests that the Lokiarchaeota (and Asgard group) are the closest relatives of Eukaryotes.

In conclusion, the Asgard group is most likely to be the close relative of Eukaryotes (Zaremba-Niedzwiedzka et al. 2017), although further studies are required to establish the details of the relationship between the Eukaryotes and the Asgard group.

Table 8.4 Distribution of “Eukaryote core genes” and some other genes

		Archaeobacteria														
		Asgard group		TACK superphylum												
		Eukarya	Lokarchaeota	Throarchaeota	Oidmarchaeota	Hinddularcheaeota	Thaumarchaeota	Algarcheota	Bathyarchaeota	Korarchaeota	Verruarchaeota	Cenarchaeota	Georcheaeota	Euryarchaeota	DPANN superphylum	
Information processing	DNA polymerase, epsilon-like	●			●											
	Topoisomerase IB	●	●			●	●	●		●						●
	RNA polymerase, A fused	●			●	●	●	●								●
	RNA polymerase, subunit G (Rpb8)	●			●	●				●		●				
	Ribosomal protein L22e	●	●		●	●				●						
	Ribosomal protein L28e/Mark16	●			●											
Cell division/ cytoskeleton	(ar)tubulins	●			●			●								
	(cren)actins	●	●	●	●	●		●	●	●	●	●				
	Gelsolin-domain protein	●	●	●	●	●		●								
	Profilin	●	●	●	●	●										
Endosomal sorting	ESCRT-I: Vps28-like	●	●		●	●										
	ESCRT-I: Steadiness box domain	●	●	●	●	●										
	ESCRT-II: VpsEAP30 domain	●	●	●	●	●										
	ESCRT-II: Vps25-like	●	●	●	●	●										
	ESCRT-III: Vps2/24/46-like	●	●	●	●	●	●	●						●		
	ESCRT-III: Vps20/32/60-like	●	●	●	●	●										
Ubiquitin system		●	●		●	●			●							
Trafficking machinery	Expansion of small GTPases	●	●	●	●	●									●	●
	Longin-domain protein	●	●	●	●	●										
	Eukaryotic RKC7 family protein	●	●	●	●	●	●									
	TRAPP-domain protein	●		●												
	Sec23/24-like protein	●		●												
	Arrestin-domain	●	●													
	WD40-Armadillo gene cluster	●		●												
OST (oligosaccharide transferase)	Ribophorin I	●	●	●	●	●	●		●							
	OST3/OST6-like	●	●	●	●	●										
	STT3-like	●	●	●	●	●										

Modified from Figure 4d of Spang et al. (2017).

●: Present in all members, ○: Present in some members, ●: Distant homologs, ●: Putative homologs.

8.4 The Host of Mitochondria and Possible Characteristics of the Proto-eukaryotic Cell

As described above, the acquisition of mitochondria was an essential event in early phase of eukaryogenesis. Three hypotheses for this process are considered here (Fig. 8.2): “serial endosymbiosis hypothesis” (Sagan 1967; Margulis 1970); “hydrogen hypothesis” (Martin and Müller 1998); and “syntrophy hypothesis” (López-García and Moreira 1999). In the serial endosymbiosis hypothesis, a

thermoplasma-like Archaeobacterium without a cell wall is considered to be the host of an alphaproteobacterial endosymbiont, the supposed ancestor of mitochondria (Margulis 1970). In the hydrogen hypothesis, symbiosis between Alphaproteobacteria and methanogens is considered to be the first step of eukaryogenesis (Martin and Müller 1998). In an anaerobic environment, Alphaproteobacteria and methanogens might have exchanged their metabolic products (H_2 and CO_2). Transition might have occurred from simple coexistence of Alphaproteobacteria and methanogens in the same environment with tight contact that enabled efficient exchange of metabolic products. Following gene transfer from Alphaproteobacteria to the host (methanogens), the alphaproteobacterial symbionts become mitochondria (and/or hydrogenosomes). In the syntrophy hypothesis, methanogens, *Deltaproteobacteria*, and Alphaproteobacteria are proposed to have contributed to eukaryogenesis (López-García and Moreira 1999). Sulfate-reducing *Deltaproteobacteria* provide hydrogen and CO_2 by fermentation, and these products are consumed by methanogens. Methanotrophic Alphaproteobacteria, candidate mitochondrial ancestors, are also thought to have contributed to ectosymbiosis. The interdependence of the three groups would then have become tighter. As a consequence, the cytoplasm and cellular membrane might have originated from *Deltaproteobacteria*, the nucleus from methanogenic Archaea, and mitochondria from methanotrophic Alphaproteobacteria.

Both the hydrogen and syntrophy hypotheses require a contribution by methane-producing Archaeobacteria. However, no members of Asgard group are known to be methanogens. Interestingly, genes for the methane synthesizing pathway were found recently in Bathyarchaeota and Verstraetearchaeota, phyla that belong to the TACK superphylum (Evans et al. 2015; Vanwonterghem et al. 2016; Borrel et al. 2016; Spang et al. 2017). Raymann et al. (2015) suggested a methanogenic euryarchaeal origin for Archaeobacteria. Therefore, the presence of methane synthesizing capacity in both the euryarchaeal branch and TACK/Asgard branch, together with the phylogenetic analysis by Raymann et al. (2015), suggests that methane production might be a symplesiomorphic (= shared ancestral) trait for archaeobacterial groups. If this suggestion is substantiated, then even early members of Asgard group might have had a methane-producing system. In addition, Lokiarchaea has been suggested to be hydrogen-dependent chemotroph (Sousa et al. 2016). If it were true, Asgard group could have been the archaeobacterial contributor to eukaryogenesis under the processes proposed by the hydrogen and syntrophy hypotheses (Sousa et al. 2016; Martin et al. 2016). By understanding the nature of Asgard archaeobacteria further, it will be clear whether these hypotheses are true or not.

8.5 Complex Process of Eukaryogenesis: Chimeric Origin of Eukaryotes

Even though the Asgard group is the strongest candidate as the archaeobacterial ancestor of Eukaryotes, the process of eukaryogenesis does not seem to be simple. When individual genes are separately analyzed, eukaryotic genes have quite diverse

origins (Table 8.3). For example, Thiergart et al. (2012) identified two major ancestors of eukaryotic genes: first, Alphaproteobacteria as the ancestor of mitochondria, and, second, Euryarchaeota as the archaeobacterial ancestor of Eukaryotes. Rochette et al. (2014) reported diverse origins of eukaryotic genes. Among genes with archaeobacterial ancestry, most were assigned to euryarchaeal, crenarchaeal, thaumarchaeal, or korarchaeal origins. Among those with bacterial ancestry, alphaproteobacterial ancestry was found and also that of other eubacterial phyla. Relatively few eukaryotic genes support the three-domain hypothesis compared with genes supporting archaeobacterial or eubacterial ancestry.

Furukawa et al. (2017) carefully analyzed individual aminoacyl tRNA synthetase (ARS) genes and found that the topologies of their phylogenetic trees varied considerably. Of 23 ARS genes analyzed, only 7–12 eukaryotic genes active in the cytoplasm seem to be of archaeobacterial origin. Only two ARS genes support the TACK/Asgard origin of Eukaryotes, though this study was carried out before the discovery of the Asgard group, and only the Lokiarchaeota were included in the data set. Instead, the PMW group (Parvarchaeota, Micrarchaeota, and Woesearchaeota) of the DPANN superphylum appeared as the closest group of Eukaryotes for six ARS genes. The DPANN superphylum includes organisms with small cell sizes and contains species that are parasitic on other Archaeobacteria (Podar et al. 2008; Rinke et al. 2013). Thus, horizontal gene transfer (HGT) from DPANN archaeobacteria to an Archaeobacterium ancestor of the Eukaryotes might have occurred because of their parasite-host relationship. Furthermore, with regard to the ARS genes of Eukaryotes that are of eubacterial origin, none are derived from Alphaproteobacteria. Various archaeobacterial genomes are known to contain genes obtained by HGT, and a range of natural transformation systems are present in Archaeobacteria (Wagner et al. 2017). Viruses (phages) or DPANN archaeobacteria might have played a role in HGT toward the Eukaryote ancestor.

The studies described here emphasize that the chimeric origin of eukaryotic genes is more complex than suggested from the acquisition of mitochondria (and plastids) by an archaeobacterial host (Furukawa et al. 2017; Fig. 8.3). As individual genes are comparatively short, it has been postulated that they do not have enough information to clarify their history. Longer sequences, formed by concatenation of genes (so-called genome-based analysis), have more phylogenetic information than individual genes. However, as various researchers have indicated, the resultant phylogeny can vary among studies using a concatenated gene approach depending on which genes are included (Da Cunha et al. 2017; Table 8.2). Thus, accumulating single gene phylogenies may be a better strategy to understand the establishment of important functions in eukaryogenesis.

8.6 Problem Regarding the Lipid Membrane Surrounding the Cell

Though all the cells of life are surrounded by lipid membrane, lipid in Archaeobacteria is totally different from those in two other domains. In eubacterial and eukaryotic cells, membrane lipids have a glycerol *sn*-3 phosphate backbone (G3P), and fatty

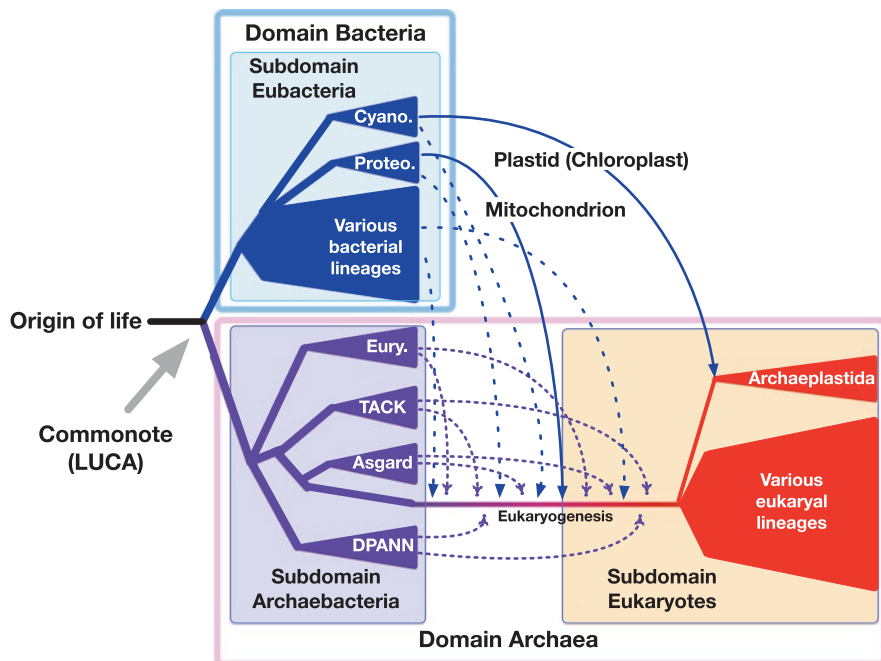


Fig. 8.3 Our view on the early phase of eukaryogenesis and relationships among cellular terrestrial life. (This figure is modified from Furukawa et al. 2017). Thick lines indicate phylogeny. Thin solid lines indicate acquisitions of mitochondrion and plastid (chloroplast). Thin dotted lines indicate horizontal gene transfer. Two domains in cellular organisms are proposed: Bacteria and Archaea. Domain Bacteria consists of single subdomain Eubacteria. Domain Archaea consists of prokaryotic subdomain Archaeobacteria and eukaryotic subdomain Eukaryotes

acid chains are connected by an ester bond. In archaeobacterial cells, lipids have a glycerol *sn*-1 phosphate (G1P) backbone, and isoprenoid chains are connected by an ether bond (See Table 8.1; see also Yokobori et al. 2016). There is no exception on the chirality of glycerol phosphate backbone, G3P or G1P, in eubacterial, eukaryotic, and archaeobacterial cellular membranes, although there are some exceptions on the distribution of ester/ether bond lipids and fatty acid/isoprenoid chains (Langworthy et al. 1983; Burggraf et al. 1992; See also Villanueva et al. 2014). G3P and G1P are synthesized from the same precursor (dihydroxyacetone phosphate). However, enzymes forming G3P (G3P dehydrogenase: G3PDH) and G1P (G1P dehydrogenase: G1PDH) have no phylogenetic and structural relationship (see Yokobori et al. 2016). In the reported genomes from Lokiarchaeota, no G1PDH gene has been found (Spang et al. 2015). However, Thorarchaeota and Heimdallarchaeota genomes have been reported to encode G1PDH gene (Zaremba-Niedzwiedzka et al. 2017). Having G1PDH gene is thought to be common ancestral characteristic of Asgard group. Therefore, during eukaryogenesis (after separating eukaryotic ancestor from Asgard group and other archaeobacteria), Archaeobacteria-type membrane was replaced with Eubacteria-type membrane.

One of the most serious problems on the transition from Archaeobacteria-type membrane to Eukaryote-type membrane seemed to be how membrane proteins working in Archaeobacteria-type membrane have adapted to the Eukaryote-type membrane; Archaeobacterial ones have isoprenoid chains, whereas eukaryotic ones have fatty chains as long hydrophobic chains. For the stability and function, interaction between membrane proteins with membrane lipids would be important. Clarifying the adaptation process of membrane proteins with archaeobacterial origin to Eukaryote-type membrane is important, since many unique characteristics of eukaryotic cells (remolding of membrane, phagocytosis, etc.) are related to the nature of membrane structure and membrane components such as membrane lipids and proteins.

8.7 Important Steps in Eukaryogenesis and Remaining Questions

Our survey of the literature highlighted three important aspects of eukaryogenesis: (1) Based on the molecular phylogenetic analysis of ribosomal protein genes and core eukaryotic genes related to cell-shape remolding and the ubiquitin system, the Asgard archaeobacterial group is the closest relative of Eukaryotes (Zaremba-Niedzwiedzka et al. 2017; see Table 8.4). (2) Acquisition of mitochondria by endosymbiosis of Alphaproteobacteria was an important step in eukaryogenesis. (3) Large-scale HGT to the Eukaryote ancestor from various lineages of Archaeobacteria and Eubacteria occurred (Furukawa et al. 2017; see also Eme et al. 2017). The HGT occurred before and after acquisition of mitochondria, although a recent study suggests acquisition of mitochondria might have occurred at a late step of eukaryogenesis (Pittis and Gabaldón 2016).

Even if we accept the major steps listed above, there are still many open questions. For example, the origin of nuclear and cellular compartmentalization in the eukaryotic cell is unclear (see Martin et al. 2015). In addition, how the archaeobacterial-type membrane was replaced by a Eubacteria-/Eukaryote-type membrane during eukaryogenesis is also uncertain (López-García and Moreira 2015). Possibly, the membrane proteins in eukaryotic cells may have been functional in the archaeobacterial membrane system during the early phase of eukaryogenesis.

8.8 Conclusion

In this chapter, we discussed the chimeric origin of Eukaryotes based on evidence from genome/phylogenetic analyses. These types of study have provided insights into the origins of components important for eukaryogenesis but do not illuminate the process of eukaryogenesis. To address how eukaryogenesis occurred, experimental approaches will be required. For example, under the concept of synthetic

biology, intermediate cells (with chimeric cellular membrane consisting of archaeobacterial membrane lipids and eubacterial/eukaryotic membrane lipids, with eukaryotic membrane-associated proteins with archaeobacterial membrane lipids, and so on) should be studied. Resurrection and characterization of ancestral proteins (and ancestral protein complexes) which are key for eukaryogenesis (such as components of complex for remodeling of cellular membrane) will also help our understanding on the process of eukaryogenesis.

References

- Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UC, Podowski RM, Näslund AK, Eriksson AS, Winkler HH, Kurkand CG (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396:133–140
- Barton NH, Briggs DEG, Eisen JA, Goldstein DB, Patel NH (2007) *Evolution*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Borrel G, Adam PS, Gribaldo S (2016) Methanogenesis and the wood–Ljungdahl pathway: an ancient, versatile, and fragile association. *Genome Biol Evol* 8:1706–1711
- Burggraf S, Olsen GJ, Stetter KO, Woese CR (1992) A phylogenetic analysis of *Aquifex pyrophilus*. *Syst Appl Microbiol* 15:352–356
- Castelle CJ, Wrighton KC, Thomas BC, Hug LA, Brown CT, Wilkins MJ, Frischkorn KR, Tringe SG, Singh A, Markillie LM, Taylor RC, Williams KH, Banfield HF (2015) Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr Biol* 25:690–701
- Cavalier-Smith T (2002) The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int J Syst Evol Microbiol* 52:7–76
- Ciccarelli FD, Doerks T, Von Mering C, Creevey CJ, Snel B, Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311:1283–1287
- Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM (2008) The archaeobacterial origin of eukaryotes. *Proc Natl Acad Sci U S A* 105:20356–20361
- Da Cunha V, Gaia M, Gadelle D, Nasir A, Forterre P (2017) Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet* 13(6):e1006810. <https://doi.org/10.1371/journal.pgen.1006810>
- Dacks JB, Field MC, Buick R, Eme L, Gribaldo S, Roger AJ, Brochier-Armanet C, Devos DP (2016) The changing view of eukaryogenesis – fossils, cells, lineages and how they all come together. *J Cell Sci* 129:3695–3703
- Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, Randau L, Hedlund BP, Brochier-Armanet C, Kunin V, Anderson I, Lapidus A, Golsman E, Barry K, Koonin EV, Hugenholtz P, Kyrpides N, Wanner G, Richardson P, Keller M, Stetter KO (2008) A korarchaeal genome reveals insights into the evolution of the archaea. *Proc Natl Acad Sci U S A* 105:8102–8107
- Eme L, Sharpe SC, Brown MW, Roger AJ (2014) On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *Cold Spring Harb Perspect Biol* 6:a016139
- Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG (2017) Archaea and the origin of eukaryotes. *Nat Rev Microbiol* 15:711–723
- Evans PN, Parks DH, Chadwick GL, Robbin SJ, Orphan VJ, Golding SD, Tyson GW (2015) Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* 350:434–438
- Foster PG, Cox CJ, Embley TM (2009) The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. *Philos Trans R Soc Lond Ser B Biol Sci* 364:2197–2207

- Fuerst JA, Sagulenko E (2011) Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function. *Nat Rev Microbiol* 9:403–413. <https://doi.org/10.1038/nrmicro2578>
- Furukawa R, Nakagawa M, Kuroyanagi T, Yokobori S, Yamagishi A (2017) Quest for ancestors of eukaryal cells based on phylogenetic analyses of aminoacyl tRNA synthetases. *J Mol Evol* 84:51–66
- Guy L, Ettema TJG (2011) The archaeal ‘TACK’ superphylum and the origin of eukaryotes. *Trends Microbiol* 19:580–587
- Harish A, Kurland CG (2017) Akaryotes and Eukaryotes are independent descendants of a universal common ancestor. *Biochimie* 138:168–183
- Harris JK, Kelley ST, Spiegelman GB, Pace NR (2003) The genetic core of the universal ancestor. *Genome Res* 13:407–412
- Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, HERNSDORF AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF (2016) A new view of the tree of life. *Nat Microbiol* 1:16048. <https://doi.org/10.1038/nmicrobiol.2016.48>
- Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T (1989) Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc Natl Acad Sci U S A* 86:9355–9359
- Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, Novák L, Zárský V, Barlow LD, Herman EK, Soukal P, Hroudová M, Doležal P, Stairs CW, Roger AJ, Eliáš M, Dacks JB, Vlček Č, Hampl V (2016) A eukaryote without a mitochondrial organelle. *Curr Biol* 26:1274–1284
- Kelly S, Wickstead B, Gull K (2011) Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. *Proc R Soc Lond B Biol Sci* 278:1009–1018
- Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, Bryant D, Hazkani-Covo E, McInerney JO, Landan G, Martin WF (2015) Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* 524:427–432
- Lake JA, Rivera MC (1994) Was the nucleus the first endosymbiont? *Proc Natl Acad Sci U S A* 91:2880–2881
- Lake JA, Henderson E, Oakes M, Clark MW (1984) Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc Natl Acad Sci USA* 81:3786–3790
- Langworthy TA, Holzer G, Zeikus JG, Tornabene TG (1983) Iso- and anteiso-branched glycerol diethers of the thermophilic anaerobe *Thermodesulfotobacterium commune*. *Syst Appl Microbiol* 4:1–17
- López-García P, Moreira D (1999) Metabolic symbiosis at the origin of eukaryotes. *Trends Biochem Sci* 24:88–93
- López-García P, Moreira D (2015) Open questions on the origin of eukaryotes. *Trends Ecol Evol* 30:697–708
- Margulis L (1970) *Origin of eukaryotic cells*. Yale University Press, New Haven
- Martin W, Müller M (1998) The hydrogen hypothesis for the first eukaryote. *Nature* 392:37–41
- Martin WF, Garg S, Zimorski V (2015) Endosymbiotic theories for eukaryote origin. *Phil Trans R Soc B* 370:20140330. <https://doi.org/10.1098/rstb.2014.0330>
- Martin WF, Neukirchen S, Zimorski V, Gould SB, Sousa FL (2016) Energy for two: new archaeal lineages and the origin of mitochondria. *BioEssays* 38:850–856
- Nelissen B, Van de Peer Y, Wilmotte A, De Wachter R (1995) An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences. *Mol Biol Evol* 12:1166–1173
- Nunoura T, Takaki Y, Kakuta Y, Nishi S, Sugahara J, Kazama H, Chee GJ, Hattori M, Kanai A, Atomi H, Takai K, Takami H (2011) Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res* 39:3204–3223

- Pisani D, Cotton JA, McInerney JO (2007) Supertrees disentangle the chimerical origin of eukaryotic genomes. *Mol Biol Evol* 24:1752–1760
- Pittis AA, Gabaldón T (2016) Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* 531:101–104
- Podar M, Anderson I, Makarova KS et al (2008) A genomic analysis of the archaeal system *Ignicoccus hospitalis*-*Nanoarchaeum equitans*. *Genome Biol* 9:1–18
- Raymann K, Brochier-Armanet C, Gribaldo S (2015) The two-domain tree of life is linked to a new root for the Archaea. *Proc Natl Acad Sci U S A* 112:6670–6675
- Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpidis NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499:431–437
- Rivera MC, Lake JA (1992) Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* 257:74–76
- Rochette NC, Brochier-Armanet C, Gouy M (2014) Phylogenomic test of the hypotheses for the evolutionary origin of eukaryotes. *Mol Biol Evol* 31:832–845
- Roger AJ, Svård SG, Tovar J, Clark CG, Smith MW, Gillin FD, Sogin ML (1998) A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proc Natl Acad Sci U S A* 95:229–234
- Sagan L (1967) On the origin of mitosing cells. *J Theor Biol* 14:255–274
- Saruhashi S, Hamada K, Miyata D, Horiike T, Shinozawa T (2008) Comprehensive analysis of the origin of eukaryotic genomes. *Genes Genet Syst* 83:285–291
- Serbus LR, Casper-Lindley C, Landmann F, Sullivan W (2008) The genetics and cell biology of *Wolbachia*-host interactions. *Annu Rev Genet* 42:683–707
- Sousa FL, Neukirchen S, Allen JF, Lane N, Martin WF (2016) Lokiarchaeon is hydrogen dependent. *Nat Microbiol* 1:16034. <https://doi.org/10.1038/nmicrobiol.2016.34>
- Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJ (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521:173–179
- Spang A, Caceres EF, Ettema TJG (2017) Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* 357:eaf3883. <https://doi.org/10.1126/science.aaf3883>
- Thiergart T, Landan G, Schenk M, Dagan T, Martin WF (2012) An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. *Genome Biol Evol* 4:466–485
- Ueno Y, Ono S, Rumble D, Maruyama S (2008) Quadruple sulfur isotope analysis of ca. 3.5 Ga Dresser Formation: new evidence for microbial sulfate reduction in the early Archean. *Geochim Cosmochim Acta* 72:5675–5691
- Vanwonterghem I, Evans PN, Parks DH, Jensen PD, Woodcroft BJ, Hugenholtz P, Tyson GW (2016) Methylophilic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat Microbiol* 1:16170. <https://doi.org/10.1038/nmPNMicrobiol.2016.170>
- Villanueva L, Sinninghe Damsté JS, Schouten S (2014) A re-evaluation of the archaeal membrane lipid biosynthetic pathway. *Nat Rev Microbiol* 12:438–448
- Wagner A, Whitaker RJ, Krause DJ, Heilers J-H, van Wolferen M, van der Does C, Albers S-V (2017) Mechanisms of gene flow in archaea. *Nat Rev Microbiol* 15:492–501
- Williams TA, Embley TM (2014) Archaeal “Dark Matter” and the origin of Eukaryotes. *Genome Biol Evol* 6(3):474–481
- Williams TA, Foster PG, Nye TM, Cox CJ, Embley TM (2012) A congruent phylogenomic signal places eukaryotes within the Archaea. *Proc R Soc Lond B Biol Sci* 279:4870–4879
- Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51:221–271
- Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. *Proc Natl Acad Sci* 87:4576–4579

- Yokobori S, Yamagishi A (2013) Birth of Eukaryotes (eukaryotic cell). In: Yamagishi A (ed) *Astrobiology: seeking origin of life in space*, Dojin Biosciences No. 6. Kagaku Dojin, Kyoto, pp 156–168 (in Japanese)
- Yokobori S, Nakajima Y, Akanuma S, Yamagishi A (2016) Birth of archaeal cells—molecular phylogenetic analyses of G1P dehydrogenase, G3P dehydrogenases, and glycerol kinase suggest derived features of archaeal membranes having G1P-polar lipids. *Archaea*. Article ID 1802675. <https://doi.org/10.1155/2016/1802675>
- Yutin N, Makarova KS, Mekhedov SL, Wolf YI, Koonin EV (2008) The deep archaeal roots of eukaryotes. *Mol Biol Evol* 25:1619–1630
- Yutin N, Wolf MY, Wolf MI, Koonin EV (2009) The origins of phagocytosis and eukaryogenesis. *Biol Direct* 4:9. <https://doi.org/10.1186/1745-6150-4-9>
- Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU, Stott MB, Nunoura T, Banfield JF, Schramm A, Baker BJ, Spang A, Etema TJG (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541:353–358

Chapter 9

Color of Photosynthetic Systems: Importance of Atmospheric Spectral Segregation Between Direct and Diffuse Radiation



Atsushi Kume

Abstract The color of photosynthetic apparatus can be used for inferring the process of evolutionary selection of photosynthetic pigments and as possible signs of life on distant habitable exoplanets. The absorption spectra of photosynthetic apparatus have close relationships with the spectra and intensity of incident radiation. Most terrestrial plants use specific light-harvesting chlorophylls and carotenoids for photosynthesis and have pale green chloroplasts. However in aquatic ecosystems, there are phototrophs with various colors having different photosynthetic pigments. Oxygenic photosynthesis uses visible light, and far-red photons are not used for this process. While some phototrophic bacteria are able to use far-red photons for their life, they do not generate O₂.

Other aspect of light is the harmful effect of light. Although efficient light absorption is important for photosynthesis, UV and excess light absorption damages photosynthetic apparatus. In terrestrial environments, portion of incident solar radiation reaches to the surface, which are called direct radiation (PAR_{dir}), while the other are optically altered by the Earth's atmosphere, scattered by the sky and clouds, which are called diffuse radiation (PAR_{diff}). The photosynthetic systems of terrestrial plants are fine-tuned to reduce the energy absorption of PAR_{dir}. The safe use of PAR_{dir} and the efficient use of PAR_{diff} are achieved in light-harvesting complexes of terrestrial plants. In addition to the type of central star, the optical properties of the atmosphere of the planet may have significant effects on the evolution of photosynthetic systems and photoreceptors.

Keywords Absorption spectra · Atmospheric optics · Chlorophylls · Photosystem · Phycobilisome

A. Kume (✉)
Faculty of Agriculture, Kyushu University, Fukuoka, Japan
e-mail: akume@agr.kyushu-u.ac.jp

9.1 Absorption Spectra of Photosynthetic Pigments

Radiant energy is transferred by photons, which are discrete bundles of electromagnetic energy that travel at the speed of light ($c = 3.0 \times 10^8 \text{ m s}^{-1}$ in vacuum). Radiant energy behaves as both particles and waves. The relationship between energy of a photon and wavelength is represented in Planck's law and can be expressed in terms of energy flux (W m^{-2}) or photon flux ($\text{mol m}^{-2} \text{ s}^{-1}$). The energy (e) of a photon flux is related to its wavelength (λ , m). For a given λ and for a mole of photons:

$$e_\lambda = N_A hc / \lambda$$

where N_A is Avogadro's number (6.022×10^{23}) and h is Planck's constant ($6.63 \times 10^{-34} \text{ J s}$). Photons with shorter wavelengths have higher energy content than those with longer wavelengths. The solar radiation spectra can be described in terms of energy (Fig. 9.1a, b) or photon flux density (Fig. 9.1c, d), each displaying different profiles.

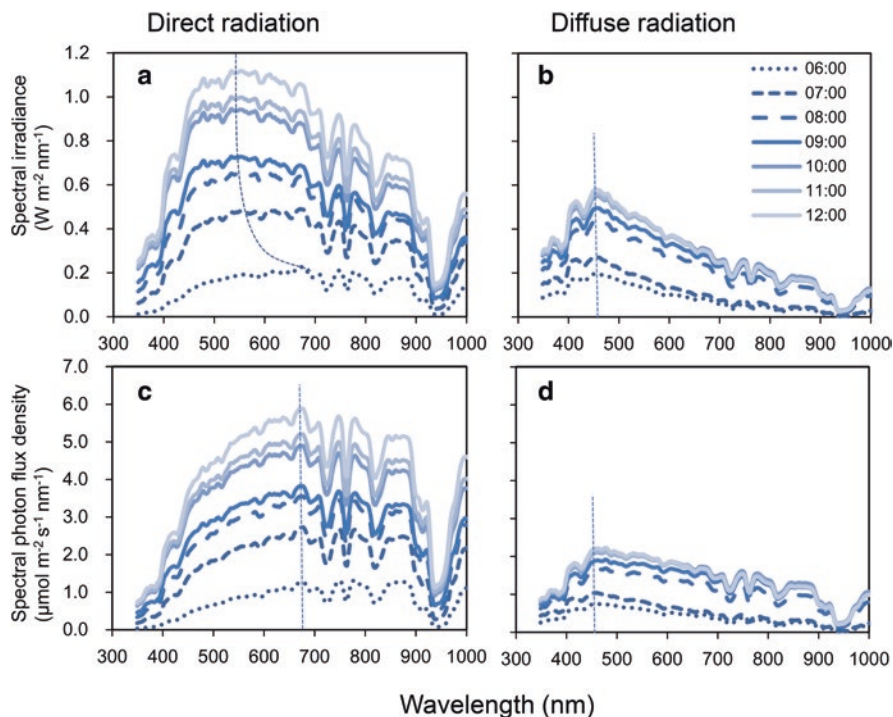


Fig. 9.1 An example of clear sky spectral irradiance and photon flux density measured by the solar tracking spectroradiometers set atop the building of the National Institute for Environmental Studies, NIES (36.05°N, 140.12°E, 40 m a.s.l.): (a, c) direct radiation and (b, d) diffuse radiation. These were measured on a clear day (day of year = 195) in 2011 and averaged over each hour. Dashed lines indicate peak wavelength (λ_{max}). (After Kume et al. 2016)

All known oxygenic photosynthesis consists of two photochemical reaction centers, photosystem I (PSI) and photosystem II (PSII) (Fig. 9.2a, b), and the excitation wavelengths for the reaction center pigments are 700 and 680 nm, respectively. Chlorophyll (Chl) *a* and Chl *b* act as key components of the photosystems, which are present as large pigment-protein assemblies in the chloroplast (Kunugi et al. 2016). The light is absorbed mainly by light-harvesting chlorophyll *a/b*-binding protein complexes (LHCs). Then the energy is transferred to PSI and PSII. Although there are several light-harvesting pigments, most terrestrial plants use specific light-harvesting chlorophylls, Chl *a* and Chl *b* (Fig. 9.3), and carotenoids, they are assembled in pigment-protein complexes (Björn et al. 2009; Kiang et al. 2007a). All Chl *b* is present in LHCs that works as peripheral-antenna-protein complexes (Fig. 9.2a, b).

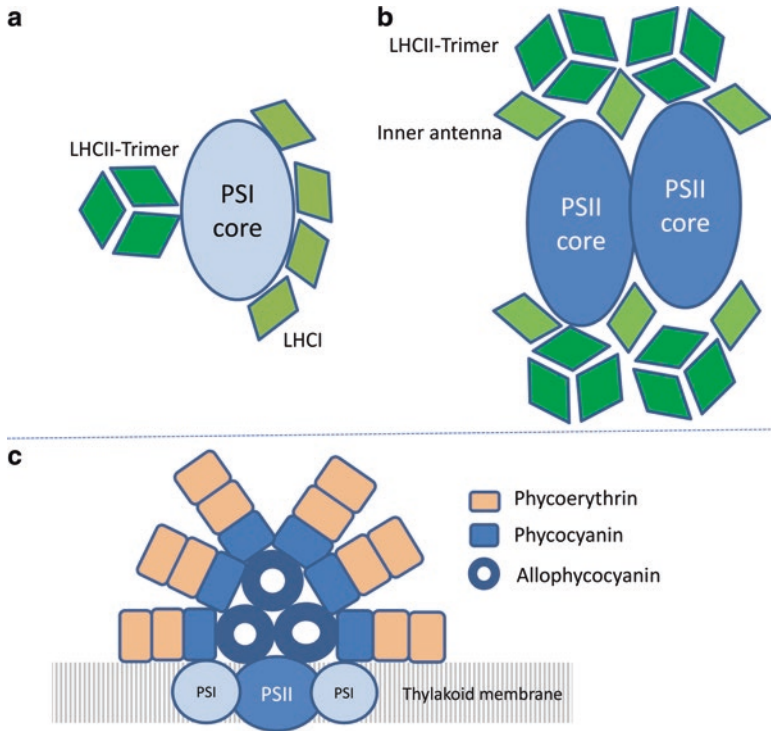


Fig. 9.2 Arrangement of the supercomplex of PSI and PSII in the thylakoid membrane. (a) PSI core is linked to the four Lhca proteins that form LHCI (pale green). LHCII also associates with PSI, in a phosphorylated process known as state transitions, to rapidly respond to blue light (Longoni et al. 2015). (b) PSII cores are linked with LHCII, consisting of Lhcb proteins (Lhcb1, Lhcb2, and Lhcb3) and inner antennas (Lhcb 4, Lhcb 5, and Lhcb6). Only the antenna proteins contain Chl *b*. (c) Simplified structural model of a phycobilisome and various phycobiliproteins, such as allophycocyanin (APC), phycocyanin (PC), and phycoerythrin (PE), which work as the antenna system. The three circles represent the tricylindrical core APC and two cylinders at the bottom attach to the thylakoid membrane. Six rods each consisting of PC and PE are attached to the three APC cores. These proteins are connected with their associated linker polypeptides (not shown). (a, b) are overhead views, and (c) is a sectional view of thylakoid membrane

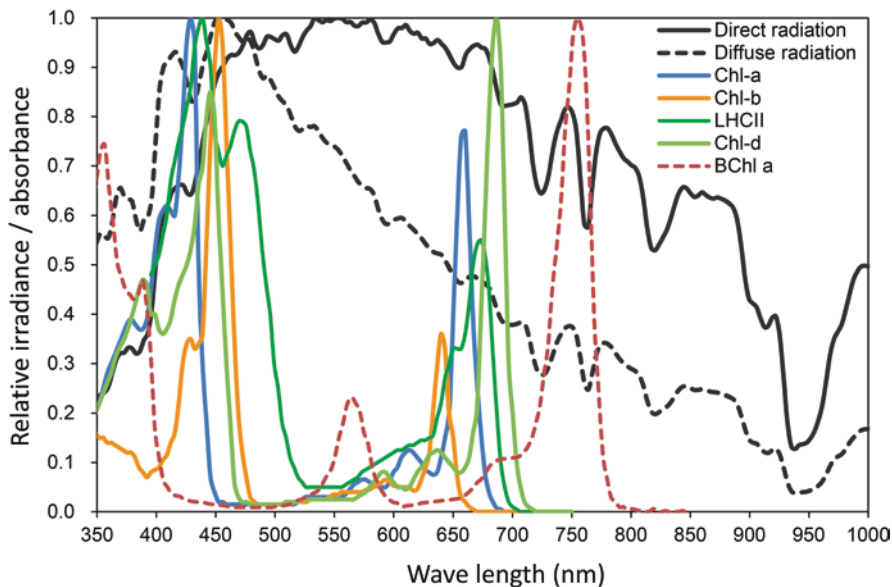


Fig. 9.3 Relative absorbance spectra of Chl *a*, Chl *b*, Chl *d*, and BChl *a* dissolved in diethyl ether (Mimuro et al. 2011), LHCII trimer (Hogewoning et al. 2012), and relative spectral irradiances of direct and diffuse solar radiation. Radiation spectra were measured at noon on a clear day (day of year = 195) in 2011

The antenna size of PSII in green plants varies depending on the amount of LHC associated with the core complexes of PSII (Jansson 1994). Spectra of pigments shift reflecting the environments in the proteinaceous environment in the complex, and pigment complexes show spectra significantly different from that of the constituting pigments (e.g., LHCII; Fig. 9.3).

The relationships between the incident radiation spectra and those of light-harvesting pigments of organisms are crucial for understanding photosynthetic life on Earth. In aquatic ecosystems, various chlorophylls (e.g., Chls *a*, *b*, *c*₁, *c*₂, *c*₃, *d*, and *f*) and increased amounts of accessory pigments, such as various carotenoids and/or phycobiliproteins (Fig. 9.2c), utilize prevailing blue-green light at depths with low light intensity (Croce and van Amerongen 2014; Kirk 2011). Phycobilisomes can absorb nearly all the available blue-green photons (Kirk 2011; Larkum 2006), and therefore, they are advantageous for photosynthesis in deep water. Chl's *d* and *f* are found in some specific cyanobacteria and are produced in response to near-infrared radiation (Gan and Byrant 2015). The absorption maxima of Chl *d* and *f* are red-shifted comparing with all the other chlorophylls and can utilize far-red light (700–750 nm), which is not absorbed by Chl *a* (Fig. 9.3; Chen and Blankenship 2011). Mielke et al. (2011) suggested that the energy conversion efficiency of Chl *d*-utilizing light reaction centers is improved by approximately 5% comparing with that of the Chl *a*-utilizing reaction centers, without decreasing the quantum efficiency.

Significantly different types of photosynthetic organisms can be found in anoxic environment, such as deep sea, hot springs, stagnant water bodies, and microbial mats in intertidal zones. They are called phototrophic bacteria. The phototrophic bacteria use bacteriochlorophylls (BChls), which mainly absorb 700–900 nm, they rely on the carotenoid and BChl absorption bands at shorter wavelengths under very deep water. Many of the phototrophic bacteria are not autotrophic; i.e., they are unable to fix CO₂ directly and require anoxic environments to thrive (Kiang et al. 2007a). The ecological importance of aerobic, anoxygenic, and phototrophic (AAP) bacteria containing BChl *a* have been revealed recently. AAP bacteria require oxygen for growth and BChl *a* synthesis and are able to use both light and organic substrates for energy production. It has been suggested that AAP bacteria play a significant role in aquatic food webs and biogeochemical cycles (Koblížek et al. 2007; Kolber et al. 2001), though the precise roles of BChl *a* for AAP bacteria remain unclear (Cottrell et al. 2010). Although the waveband active for oxygenic photosynthesis is limited to around 400–700 nm, that for anoxygenic photosynthesis using BChl extends up to 1000 nm or more (Kiang et al. 2007a). Therefore, the absorption spectrum of BChl *a* does not overlap with that of Chl *a* and Chl *b* (Fig. 9.3), and AAP bacteria are able to absorb the light transmitted through vegetation. It is expected that several new ecological niches of AAP bacteria will be found, including those within the terrestrial environments, such as shaded phytotelmata: water-filled cavities in a terrestrial plant (Lehours et al. 2016).

9.2 Possibility of Multi-photosystems

Terrestrial oxygenic photosynthesis consists of tandemly connected two photochemical reaction centers, PSI and PSII, with two-step linear electron flow (see Chap. 10). The energy required for CO₂ reduction and O₂ evolution is supplied by two photosystems absorbing light at 700–730 nm. On other habitable exoplanets, for example, those around M-dwarf star having lower surface temperature (2500–3800 °K) than that of the Sun (5772 °K) (Kiang et al. 2007b; Walfencroft and Raven 2002), the emission spectra are peaking at near-infrared region. The energy supplied by near infrared is not sufficient to reduce CO₂ producing O₂ with the two-step photosynthesis. However, instead of the two-photon reaction, three photosystems utilizing the photons at 1050–1095 nm or four photosystems absorbing photons of 1400–1460 nm are proposed to promote the oxygenic photosynthesis (Kiang et al. 2007b; Walfencroft and Raven 2002). The redox power of the multiphoton reactions may be higher than the two-photon reaction. Takizawa et al. (2017) discussed the possible adaptive evolution of land phototrophs from marine phototrophs on a hypothetical habitable planet around an M-dwarf star and examined “two-color” reaction centers using photosynthetically active radiation (PAR, 400–700 nm) and near-infrared radiation (NIR, 700–1400 nm). They suggested the possible evolutionary process of land phototrophs with multi-photosystems. They have noted that the illumination spectra under water of the planet shift from NIR to PAR due to

the absorption of NIR by water. Therefore, Earth-type two photosystems may be rather evolutionally advantageous underwater of M-dwarf planet.

The spectral characteristics of photosynthetic organism have to include the spectral environment in situ not only the spectral characteristics of central star.

9.3 Photodamage of Photosynthetic System

Chl *a*, *b*, and *d* and BChls show strong absorption only in the 330–480 nm and 630–1050 nm ranges (Scheer 2003). Photons of wavelength longer than 1100 nm behave as thermal energy and difficult to excite electron energy state (Kiang et al. 2007b). Ultraviolet (UV) radiation has wavelengths of 10–400 nm, i.e., shorter than PAR. In addition to the effect of photon on photosynthesis, we have to consider the effect of photons on biological materials especially DNA. UV-B radiation (280–315 nm) is known to induce highest damages on DNA because of its absorption peaks at the wavelength region (Rozema et al. 2002; Hidema et al. 2007; Wang et al. 2014). Terrestrial organisms are protected from UV radiation by atmospheric ozone. During prolonged exposure to solar or artificial UV-B, the formation of reactive oxygen species (ROS) is induced, and proteins, nucleic acids, and photosynthetic pigments are directly damaged. Mulkidjanian and Junge (1997) hypothesized that UV-screening proteins became those transfer the excitation energy to porphyrin, and then later they functioned as reaction centers and antenna proteins. Moreover, recent research on LHC-like proteins suggested that the ancestor of these proteins played photoprotective roles within the thylakoids, instead of performing photosynthesis (Ballottari et al. 2012).

In the terrestrial environment, wavelengths longer than 700 nm or shorter than 400 nm cannot be used for photosynthesis, but far-red light may have supportive functions to prevent PSI photoinhibition (Kono et al. 2017). The radiation within the 400–700 nm waveband, in the visible radiation (400–760 nm), is defined as PAR for terrestrial plants (McCree 1972). Chlorophylls do not absorb photons in the PAR waveband evenly (Fig. 9.3), only a small fraction of absorbance occurs in the green region (500–600 nm), though the photosynthetic quantum yields are similar from green to red light (Hogewoning et al. 2012). Terrestrial plants have developed blue and green photon-screening pigments, such as carotenoids (Kume et al. 2016) and anthocyanins (van den Berg et al. 2009; Merzlyak et al. 2008), to avoid photoinhibition or photooxidation, instead of light-harvesting phycobilisomes in aquatic photosynthetic organisms (Fig. 9.2).

The spectral absorbance of carotenoids is effective in eliminating shortwave PAR (e.g., β -carotene efficiently absorbs 410–490 nm), which contains much of the surplus energy that is not used for photosynthesis and is dissipated as heat (Kume 2017). The energy transfer efficiency of β -carotene is approximately 35% (de Weerd et al. 2003) and prevents excess energy inflow into the light-harvesting complexes.

Additionally, absorbed excess photon energy became the chlorophyll electron energy excited triplet state, which is the state with two electrons of the same spin in an excited energy state. Chlorophyll in triplet state is harmful for photosynthetic system. Carotenoids are efficient quenchers of chlorophyll triplet states when they are in direct contact with the chlorophylls (Krieger-Liszkay et al. 2008), and carotenoids perform an essential photoprotective role within the chloroplast (Johnson et al. 2011; Young 1991).

9.4 Spectral Characteristics Depending on Plant Leaf Morphology and Atmosphere

Absorbance is not the sole factor determining the spectral characteristics of the intact plant leaves and its efficiency. Leaf tissue structure affects the absorption properties of leaves in addition to the simple absorption properties of constituting pigments (Moss and Loomis 1952; Vogelmann 1993; Kume 2017). The photon absorption of a whole leaf depends significantly on a combination of pigment density distribution and leaf anatomical structures (Fig. 9.4) (Knipling 1970). In terrestrial plants, leaf anatomical structure and optimal chloroplast distribution enable leaves to become gray bodies for PAR and to improve PAR-use efficiency (Kume 2017).

Atmospheric CO₂ concentration also has great effects on leaf shape and transpiration. Beerling et al. (2001) demonstrated that the evolution of megaphyll, spread flat, leaves from needleleaf occurred during the geological time with a massive decrease in atmospheric CO₂ concentration in the Late Paleozoic era. The plants in a high-CO₂ environment have fewer stomata, which causes less heat dispersion through evaporation. Consequently, at high light intensities, leaves with a broad lamina are more likely to suffer from heat stress because less heat is lost through transpiration. Large leaves have much higher leaf temperatures than small leaves when stomata are closed. However, small leaves can maintain its temperature similar to the air temperature with the difference of few degrees, regardless of the stomatal conductance (Campbell and Norman 1998). The evolution of megaphylls with more transpiration was promoted after the atmospheric CO₂ concentration became lower, as the by-product of the increase of stomata. Gates et al. (1965) also pointed out the importance of low NIR absorption and high far-infrared emittance of megaphyll to avoid overheating. The light absorbance spectra of pigments and their density distribution within a leaf determine heat absorption of the leaf. This explains how the leaf color, shape, and anatomical structure are linked. The atmospheric composition, solar spectrum, and intensity on the planet biophysically determine the evolution of absorption spectra of phototrophs. Therefore, it is necessary to infer the color of the photosynthetic system of an exoplanet from these factors in addition to the incident radiation at the top of its atmosphere.

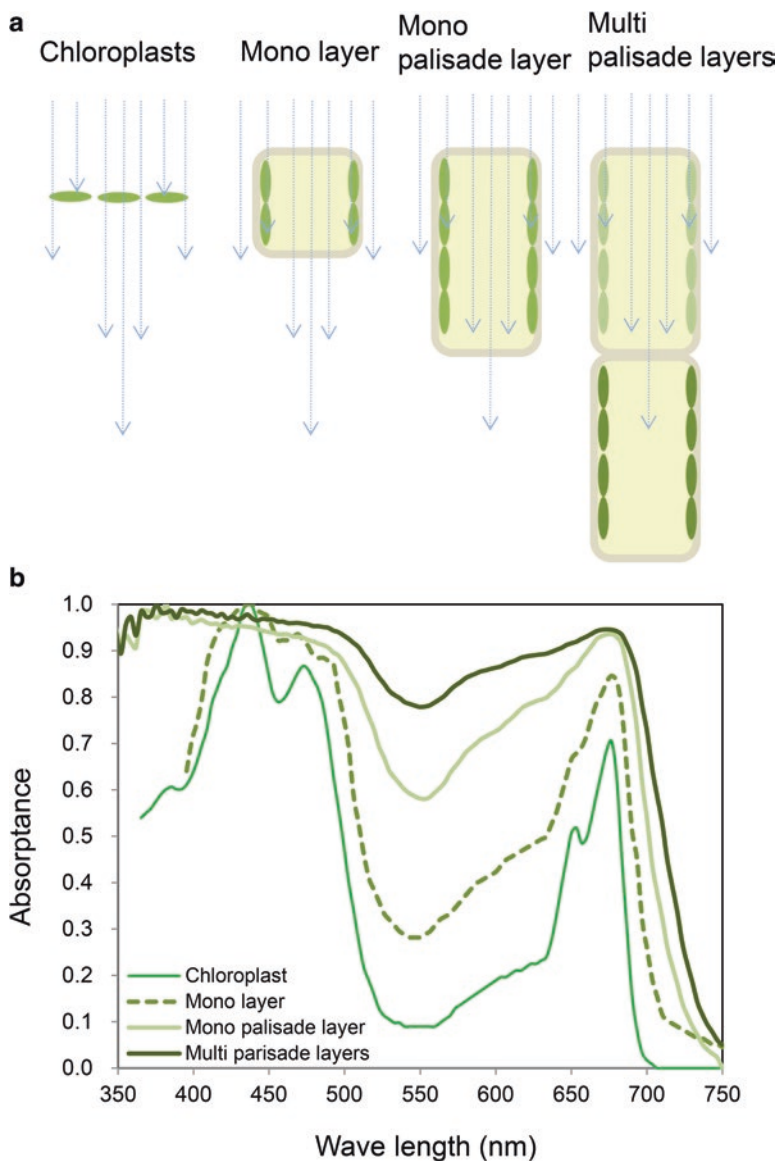


Fig. 9.4 Schematic models explaining the absorption of palisade cells (a) and absorption spectra with the same color of photosynthetic pigments (b). The absorption spectrum of a single layer of chloroplasts is similar to that of the LHCII trimer of the same concentration. The effective absorption area per unit leaf area increases with the aspect ratio, height per width, of the cell and the number of cell layers. Therefore, the absorption of incident photons per chloroplast is inversely related to the aspect ratio. In addition, the light incident on a leaf, particularly green light, is scattered by chloroplasts in leaves with well-arrayed palisade cells. The stacked palisade cells have the different absorbance between upper and lower layer (Multi palisade layer) and can absorb photons more efficiently with multiple absorptions within leaf tissues, and the absorption spectrum becomes flat. The spectra were normalized to the maxima of the Soret bands. Additional explanation can be found in Kume (2017)

9.5 Effect of Atmospheric Scattering of Solar Radiation on Incident Radiation

The function of absorption spectra has to be considered in the combination with the spectra of incident radiation (Kiang et al. 2007a; Kume et al. 2016; Takizawa et al. 2017; Wolstencroft and Raven 2002). Incident global radiation consists of two main components, direct radiation (PAR_{dir}), which arrives after transmission through the atmosphere, and diffuse radiation, which is the sunlight scattered by molecules, aerosols, and clouds (PAR_{diff}). These components are characterized by large differences in light quantity, directional characteristics, and spectral quality, and they depend on the cloud coverage and various other conditions at the location (Akitsu et al. 2015). However, the evaluation of the spectral effects of solar radiation on photosynthesis has been conducted using averaged spectra, and the effects of differences in the spectra between PAR_{dir} and PAR_{diff} have been ignored and not evaluated (e.g., Chen and Blankenship 2011; Wolstencroft and Raven 2002).

As shown in Fig. 9.1, both energy and photon flux spectra differ in magnitude and profile between the PAR classes. In photon flux of PAR_{dir} is higher in the 650–700 nm bandwidth, while the flux of PAR_{diff} is higher in the 450–500 nm bandwidth on sunny days (Fig 9.1c, d). Strong spectral irradiance of PAR_{dir} in the 500–600 nm bandwidth arise at noon on sunny days (Fig. 9.1a). As discussed in the next section, these spectral differences between PAR_{dir} and PAR_{diff} are large enough to drive adaptive selection of absorption characteristics of photosynthetic absorbers.

9.6 Effect of Strong Solar Radiation on the Spectral Absorption of Photosynthetic Organisms

Because the efficiency in conversion of light energy is important to understand biomass production, several leaf photosynthesis models that take the light absorption profile into account have been proposed. These models are based on the optimal use of PAR photons in the terrestrial ecosystem. Most of the discussion has been concentrated on the efficient use of incident PAR photons in photosynthesis. In the discussion of exoplanets, maximal photon absorption and photon-use efficiency have been the main subject (Kiang et al. 2007b; Takizawa et al. 2016; Wolstencroft and Raven 2002). As a result, the light absorption efficiency was regarded as the highest in the waveband ranges of strongest energy or abundant photon densities. However, the balance between light absorption and CO_2 assimilation of chloroplasts is crucially important for preventing ROS generation. The spectral characteristics of incident solar radiation have a significant biophysical effect on the adaptation of chloroplasts and leaf structure for preventing excess energy absorption (Kume 2017).

Preventing excess energy absorption in chloroplasts is a primal survival strategy in terrestrial environments, because typical maximum rates of net photosynthesis of $0.5\text{--}2.0 \text{ mg CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ correspond to the net heat stored in endergonic biochemical

reactions of photosynthesis, i.e., between 8 and 32 W m⁻² (Jones 2014). This indicates that 97% of incident solar energy (approximately 1000 W m⁻²) must be emitted or transmitted outside safely without causing damage or physiological inhibition. The waveband of the green region (500–570 nm) is identical to that of strong directional solar irradiance during midday under clear skies (Figs. 9.1a and 9.3), and most of the energy cannot be used for photosynthesis and must be eliminated safely without absorption. As can be seen in Fig. 9.5, absorption sharply drops in the spectral region where the spectral irradiance is high. There are clear negative relationships between the spectral absorbance of the antenna systems (PSI-LHCI and LHCII) and the spectral irradiance of PAR_{dir} at noon in the high spectral irradiance waveband (450–650 nm) (Fig. 9.5). Therefore, the antenna systems can alleviate the heat stress of strong direct solar radiation during midday (Kume et al. 2016).

It is noted that the spectral absorbance of pure Chl *a* solution decreased with the increased spectral irradiance of global PAR at noon ($r = -0.87$), and it was sufficient to avoid strong radiation energy waveband rather than absorb photon flux efficiently (Kume et al. 2016). On the other hand, the spectral absorbance of Chl *b* is unable to avoid strong PAR_{dir} radiation at noon, but it can use PAR_{diff} photons efficiently, because the Soret wavelength of Chl *b* (blue absorption band peaking approximately 452 nm in diethyl ether) is the longest among the chlorophyll pigments (Mimuro et al. 2011), enabling high absorption efficiency for PAR_{diff}. When plants are grown under low light intensities, Chl *b* synthesis is enhanced, and antenna size increases (Bailey et al. 2001; Tanaka and Tanaka 2011). The increased antennas improve the light absorption efficiency, but it may become vulnerable to the strong PAR_{dir} (Kume et al. 2018).

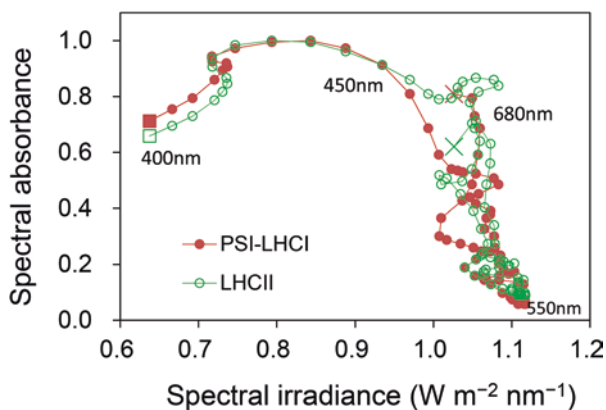


Fig. 9.5 Relationships between spectral irradiance of PAR_{dir} at noon and spectral absorbance of purified LHCII trimer and PSI-LHCI (after Kume et al. 2016). The graphs are plotted with spectral absorbance on the y-axis and the spectral irradiance on the x-axis at 3.35 nm intervals in the 400–680 nm bandwidth. Points with consecutive wavelengths are connected with a line. The points with the shortest (400 nm) and longest wavelengths (680 nm) are indicated by squares and crosses, respectively

The spectral absorbance of phycobilisome-type antennas cannot avoid direct solar radiation on land. Terrestrial green plants are fine-tuned to reduce excess energy absorption by photosynthetic pigments instead of efficient absorption of PAR photons (Kunugi et al. 2016; Kume et al. 2018). Leaves adapted to high light levels have multiple palisade layers and utilize PAR_{dir} better than PAR_{diff} (Brodersen et al. 2008). Such an anatomical structure is the key feature for efficient and safe use of direct solar radiation by the leaves of terrestrial plants (Fig. 9.4).

In addition, PAR_{dir} and PAR_{diff} may influence human vision (Purkinje shift). The light range to which humans and other animals are most sensitive (550–600 nm) falls within the range of the highest spectral radiant energy, as can be seen in LHCII. However, under low light conditions, rhodopsin, the pigment present in rod cells that mediate vision in dim light, absorbs maximally around 500 nm (Rister and Desplan 2011), where PAR_{diff} is high. Clearly, the spectral differences between sunlit and shaded areas are linked to animal vision. Overall, the spectral differences between PAR_{dir} and PAR_{diff}, as well as the steady λ_{\max} of PAR_{diff}, exert multiple effects on terrestrial organisms and must be effective drivers of diversification in pigment distribution and function.

9.7 Conclusion

Several light-harvesting pigments cover the PAR and NIR ranges of light wavelengths. These pigments are selected for several reasons: (1) effective light absorption for abstracting electrons from electron donors, (2) spectral limitations of excess absorption for protecting photo-absorbers, and (3) oxygenic or anoxygenic photosynthesis. Most terrestrial plants use specific light-harvesting chlorophylls, Chl *a* and Chl *b*, and carotenoids. The spectra of these pigments and constructed photosystems and antenna proteins significantly align with the spectra of PAR_{dir} and PAR_{diff} for the safe and efficient use of solar radiation on land. The atmospheric environment of the planet, in addition to the type of central star, may have significant effects on the evolution of photosynthetic systems and photoreceptors of organisms.

References

- Akitsu T, Kume A, Hirose Y, Ijima O, Nasahara KN (2015) On the stability of radiometric ratios of photosynthetically active radiation to global solar radiation in Tsukuba, Japan. *Agric For Meteorol* 209–210:59–68
- Bailey S, Walters RG, Jansson S, Horton P (2001) Acclimation of *Arabidopsis thaliana* to the light environment: the existence of separate low light and high light responses. *Planta* 213:794–801
- Ballottari M, Girardon J, Dall’Osto L, Bassi R (2012) Evolution and functional properties of photosystem II light harvesting complexes in eukaryotes. *BBA-Bioenergetics* 1817:143–157
- Beerling DJ, Osborne CP, Chaloner WG (2001) Evolution of leaf-form in land plants linked to atmospheric CO₂ decline in the Late Palaeozoic era. *Nature* 410:352–354

- Björn LO, Papageorgiou GC, Blankenship RE, Govindjee (2009) A viewpoint: why chlorophyll A? *Photosynth Res* 99:85–98
- Brodersen CR, Vogelmann TC, Williams WE, Gorton HL (2008) A new paradigm in leaf-level photosynthesis: direct and diffuse lights are not equal. *Plant Cell Environ* 31:159–164
- Campbell GS, Norman JM (1998) An introduction to environmental biophysics, 2nd edn. Springer, New York
- Chen M, Blankenship RE (2011) Expanding the solar spectrum used by photosynthesis. *Trends Plant Sci* 16:427–431
- Cottrell MT, Ras J, Kirchman DL (2010) Bacteriochlorophyll and community structure of aerobic anoxygenic phototrophic bacteria in a particle-rich estuary. *ISME J* 4:945–954
- Croce R, van Amerongen H (2014) Natural strategies for photosynthetic light harvesting. *Nat Chem Biol* 10:492–501. <https://doi.org/10.1038/nchembio.1555>
- De Weerd FL, Dekker JP, van Grondelle P (2003) Dynamics of β -carotene-to-chlorophyll singlet energy transfer in the core of photosystem II. *J Phys Chem B* 107:6214–6220
- Gan F, Bryant D (2015) Adaptive and acclimative responses of cyanobacteria to far-red light. *Environ Microbiol* 17:3450–3465
- Gates DM, Keegan HJ, Schleter JC, Weidner VR (1965) Spectral properties of plants. *Appl Opt* 4:11–20
- Hidema J, Taguchi T, Ono T, Teranishi M, Yamamoto K, Kumagai T (2007) Increase in CPD photolyase activity functions effectively to prevent growth inhibition caused by UVB radiation. *Plant J* 50:70–79
- Hogewoning SW, Wientjes E, Douwstra P, Trouwborst G, van Ieperen W, Croce R, Harbinson J (2012) Photosynthetic quantum yield dynamics: from photosystems to leaves. *Plant Cell* 24:1921–1935
- Jansson S (1994) The light-harvesting chlorophyll *a/b*-binding proteins. *Biochim Biophys Acta* 1184:1–19
- Johnson MP, Goral TK, Duffy CD, Brain AP, Mullineaux CW, Ruban AV (2011) Photoprotective energy dissipation involves the reorganization of photosystem II light-harvesting complexes in the grana membranes of spinach chloroplasts. *Plant Cell* 23:1468–1479
- Jones HG (2014) Plants and microclimate: a quantitative approach to environmental plant physiology, 3rd edn. Cambridge University Press, Cambridge, UK, p 407
- Kiang NY, Segura A, Tinetti G, Govindjee, Blankenship RE, Cohen M, Siefert J, Crisp D, Meadows VS (2007a) Spectral signatures of photosynthesis. II. Coevolution with other stars and the atmosphere on extrasolar worlds. *Astrobiology* 7:252–274
- Kiang NY, Siefert J, Govindjee, Blankenship RE (2007b) Spectral signatures of photosynthesis. I. Review of Earth organisms. *Astrobiology* 7:222–251
- Kirk JTO (2011) Light and photosynthesis in aquatic ecosystems. Cambridge University Press, Cambridge, UK
- Knipling EB (1970) Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sens Environ* 1:155–159
- Koblížek M, Masín M, Ras J, Poulton AJ, Práasil O (2007) Rapid growth rates of aerobic anoxygenic phototrophs in the ocean. *Environ Microbiol* 9:2401–2406
- Kolber ZS, Plumley FG, Lang AS, Beatty JT, Blankenship RE, VanDover CL, Vetriani C, Koblížek M, Rathgeber C, Falkowski PG (2001) Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science* 292:2492–2495
- Kono M, Yamori W, Suzuki Y, Terashima I (2017) Photoprotection of PSI by far-red light against the fluctuating light-induced photoinhibition in *Arabidopsis thaliana* and field-grown plants. *Plant Cell Physiol* 58:35–45
- Krieger-Liszak A, Fufezan C, Trebst A (2008) Singlet oxygen production in photosystem II and related protection mechanism. *Photosynth Res* 98:551–564
- Kume A (2017) Importance of the green color, absorption gradient, and spectral absorption of chloroplasts for the radiative energy balance of leaves. *J Plant Res* 130:501–514
- Kume A, Akitsu T, Nasahara KN (2016) Leaf color is fine-tuned on the solar spectra to avoid strand direct solar radiation. *J Plant Res* 129:615–624

- Kume A, Akitsu T, Nasahara KN (2018) Why is chlorophyll b only used in light-harvesting systems? *J Plant Res* 131(6):961–972
- Kunugi M, Satoh S, Ihara K, Shibata K, Yamagishi Y, Kogame K, Obokata J, Takabayashi A, Tanaka A (2016) Evolution of green plants accompanied changes in light-harvesting systems. *Plant Cell Physiol* 57:1231–1243
- Larkum AWD (2006) The evolution of chlorophylls and photosynthesis. In: Grimm B, Porra RJ, Rüdiger W, Scheer H (eds) *Chlorophylls and bacteriochlorophylls: biochemistry, biophysics, functions and applications, advances in photosynthesis and respiration*, vol 25. Springer, New York, pp 261–282
- Lehours AC, Jeune AL, Aguer JP, Céréghino R, Corbara B, Kéraval B1, Leroy C, Perrière F, Jeanthon C, Carrias JF (2016) Unexpectedly high bacteriochlorophyll *a* concentrations in neotropical tank bromeliads. *Environ Microbiol Rep* 6:689–698. <https://doi.org/10.1111/1758-2229.12426>
- Longoni P, Douchi D, Cariti F, Fucile G, Goldschmidt-Clermont M (2015) Phosphorylation of the light-harvesting complex II isoform Lhcb2 is central to state transition. *Plant Physiol* 169:2874–2883
- McCree KJ (1972) The action spectrum, absorptance and quantum yield of photosynthesis in crop plants. *Agric Meteorol* 9:90–98
- Merzlyak MN, Chivkunova OB, Solovchenko AE, Razi Naqvi K (2008) Light absorption by anthocyanins in juvenile, stressed, and senescing leaves. *J Exp Bot* 59:3903–3911
- Mielke SP, Kiang NY, Blankenship RE, Gunner MR, Mauzerall D (2011) Efficiency of photosynthesis in a Chl *d*-utilizing cyanobacterium is comparable to or higher than that in Chl *a*-utilizing oxygenic species. *Biochim Biophys Acta* 1807:1231–1236
- Mimuro M, Kakitani K, Tamiaki H (2011) *Chlorophylls-structure, reaction and function*. Shokabo, Tokyo, p 305
- Moss RA, Loomis WE (1952) Absorption spectra of leaves. 1. The visible spectrum. *Plant Physiol* 27:370–391
- Mulkiyanian AY, Junge W (1997) On the origin of photosynthesis as inferred from sequence analysis: a primordial UV-protector as common ancestor of reaction centers and antenna proteins. *Photosynth Res* 51:27–42
- Rister J, Desplan C (2011) The retinal mosaics of opsin expression in invertebrates and vertebrates. *Dev Neurobiol* 71:1212–1226
- Rozema J, Björn LO, Bornman JF, Gaberšček A, Häder DP, Trošt T, Germ M, Klisch M, Gröniger A, Sinha P, Lebert M, He YY, Buffoni-Hall R, de Bakker NVJ, van de Staaij J, Meijkamp BB (2002b) The role of UV-B adiation in aquatic and terrestrial ecosystems—an experimental and functional analysis of the evolution of UV-absorbing compounds. *J Photochem Photobiol B Biol* 66:2–12. [https://doi.org/10.1016/S1011-1344\(01\)00269-X](https://doi.org/10.1016/S1011-1344(01)00269-X)
- Scheer H (2003) The pigments. In: Green B, Parson WW (eds) *Light-harvesting antennas in photosynthesis, advances in photosynthesis and respiration*, vol 13. Kluwer Academic Publishers, Dordrecht, pp 29–81
- Takizawa K, Minagawa J, Tamura M, Kusakabe N, Narita N (2017) Red-edge position of habitable exoplanets around M-dwarfs. *Sci Rep* 7:7561. <https://doi.org/10.1038/s41598-017-07948-5>
- Tanaka R, Tanaka A (2011) Chlorophyll cycle regulates the construction and destruction of the light-harvesting complexes. *Biochim Biophys Acta* 1807:968–976
- van den Berg AK, Vogelmann TC, Perkins TD (2009) Anthocyanin influence on light absorption within juvenile and senescing sugar maple leaves – do anthocyanins function as photoprotective visible light screens? *Funct Plant Biol* 36:793–800
- Vogelmann TC (1993) Plant tissue optics. *Annu Rev Plant Physiol Plant Mol Biol* 44:231–251
- Wang QW, Hidema J, Hikosaka K (2014) Is UV-induced DNA damage greater at higher elevation? *Am J Bot* 101:796–802
- Wolstencroft RD, Raven JA (2002) Photosynthesis: likelihood of occurrence and possibility of detection on earth-like planets. *Icarus* 157:535–548
- Young AJ (1991) The photoprotective role of carotenoids in higher plants. *Physiol Plant* 83:702–708

Chapter 10

Evolution of Photosynthetic System



Satoshi Hanada

Abstract Cyanobacteria and chloroplasts in plants and algae possess two different light-driven engines, designated as photosystem I (PS I) and II (PS II). Each photosystem contains chlorophylls as a photosynthetic pigment and has a principal importance in the photosynthetic electron transport system. They photooxidize water as an electron donor, and oxygen is evolved as a result, which is called oxygenic photosynthesis. In the living world, however, there is another type of photosynthesis without evolving oxygen, i.e., anoxygenic photosynthesis. The anoxygenic phototrophs cannot use water as an electron donor but use various reductive compounds such as hydrogen sulfide and hydrogen instead of water. Although oxygenic photosynthesis includes two photosystems, PS I and PS II, anoxygenic phototrophs have either one of the photosystems. Anoxygenic phototrophs are widely distributed among the bacteria, whereas oxygenic photosynthesis is limited to the cyanobacterial lineage. The phylogenetic analysis strongly suggests that oxygenic photosynthesis has emerged from anoxygenic photosynthesis. Before emergence of oxygenic photosynthesis, ancestral PS I and PS II have evolved in anoxygenic phototrophs. Emergence of oxygenic photosynthesis has a close relation to the coexistence of these different photosystems in a cyanobacterial ancestor 2.5 G years ago. The coexistence occurred by lateral gene transfer (LGT), such a LGT was frequently found in the evolutionary process of anoxygenic photosynthesis. The frequent LGT of photosystems formed the phylogenetic divergence of anoxygenic phototrophs and contributed the emergence of oxygenic photosynthesis.

Keywords Oxygenic photosynthesis · Anoxygenic photosynthesis · Photosystem · Cyanobacteria · Evolution of photosynthesis

S. Hanada (✉)

Graduate School of Science, Tokyo Metropolitan University, Tokyo, Japan
e-mail: satohana@tmu.ac.jp

10.1 Introduction

The habitable zone is defined as a range of orbits around a star; there can be liquid water and atmospheric pressure suitable for life. Habitable zone is also called the Goldilocks zone, which is named after a little girl in a tale who preferred porridge “just at the right temperature,” neither too hot nor too cold. When a planet is too close to its central star, water on the planet is completely evaporated. Conversely, the whole planet is frozen eternally if the orbit is too large. The Earth is situated at “the just right position” in the solar system and is moderately warmed up by radiation from the Sun, just like Goldilocks chose porridge “just at the right temperature.” As a result, all living organisms on Earth can thrive vigorously.

A large amount of radiation from the Sun is able not only to warm up Earth but also provides enough energy to support all life in this planet. Plants and algae can use the sunlight by photosynthesis. They convert the radiant energy to bioavailable energy and make organic compounds from atmospheric carbon dioxides for their growth and support all animal life on Earth including human beings. Also, they supply a great volume of oxygen so that all heterotrophic organisms can live using it by respiration. In general, the sunlight is obviously indispensable for all life form on Earth no matter how an organism grows phototrophically or heterotrophically (see exception in Chap. 20).

Sunlight is no doubt the sole influx energy into the closed environment on Earth. Making use of this precious influx is, thus, quite advantageous to the whole telluric life and lead them to the explosive enrichment of their biomass. It would have been a great advantage for all life on Earth that some constituents could obtain ability to use sunlight, i.e., photosynthesis, whichever by chance or by necessity. Photosynthesis brought great prosperity and surprising biological diversity to the telluric ecosystem as a consequence.

Photosynthesis found in plants and algae is termed oxygenic photosynthesis since it produces oxygen. However, oxygenic photosynthesis has not been invented by these eukaryotes. Even before they emerged on Earth, an ancient prokaryotic organism akin to cyanobacteria made the great discovery 2.7 giga years ago (Gya) (Des Marais 2000). After the emergence of the first oxygenic photosynthetic bacterium, a large amount of oxygen has been discharged, and the atmospheric oxygen tension gradually increased to 100th of the current level within approximately 1 G year (Lyons et al. 2014; Xiong and Bauer 2002). Photosynthetic eukaryotes, e.g., plants and algae, which are thought to have appeared on Earth 1 Gya or later, have acquired their ability of oxygenic photosynthesis by capturing the ancient cyanobacterium and retaining it within a cell as a symbiont. An organelle for operating photosynthesis called a chloroplast originated from the symbiotic cyanobacterium (see Chap. 8).

Oxygenic photosynthesis, however, did not suddenly appear in the ancient ecosystem. Prior to emergence of cyanobacterial oxygenic photosynthesis, another type of photochemical reaction already existed. This ancient type is a photosynthetic reaction without oxygen production and is called anoxygenic photosynthesis. Oxygenic photosynthesis found in plants, algae, and cyanobacteria shares common

features in respect to their structure, pigments, and photochemical electron transport system. On the other hand, there is clear diversity within anoxygenic photosynthesis: their photosynthetic pigments and photochemical electron transport systems have a great variety. Furthermore, ability of anoxygenic photosynthesis is widely distributed in six phyla of the domain *Bacteria*, although oxygenic photosynthesis is found only in the phylum *Cyanobacteria*.

These findings mentioned above strongly suggest that oxygenic photosynthesis evolved from anoxygenic photosynthesis and various types of anoxygenic photosynthesis have evolved before appearance of the oxygenic photosynthesis (Blankenship 1992; Xiong and Bauer 2002). However, it is still controversial how oxygenic photosynthesis was established and evolved. In this chapter, I will explain early evolution of photosynthetic system that occurred 3.5–2.7 Gya, especially evolution of photochemical reaction centers that are core protein complexes essential for photosynthesis. Lateral gene transfer, the movement of genetic information between different species, is closely related to the evolution of reaction centers, and it is of a great importance to interpret evolutionary scheme of anoxygenic photosynthesis as well as the origin of oxygenic photosynthesis.

10.2 Photosynthetic Electron Transport Chain in Oxygenic Photosynthesis

Oxygenic photosynthetic cyanobacteria and chloroplasts in plants and algae possess two different types of photochemical reaction centers, i.e., photosystem I (PS I) and photosystem II (PS II), in their photosynthetic electron transport chain (Blankenship 1992). These two reaction centers are membrane-spanning protein complexes that contain chlorophyll (Chl) molecules, and each has an essential role in the photosynthetic electron transportation as an engine to drive the electron transport system (Fig. 10.1).

One of the photochemical reaction centers, PS II, is able to oxidize water with light energy. When PS II is excited by four photons, this complex pulls four electrons out of two molecules of water, resulting in the evolution of one oxygen molecule. The oxygen evolution occurs in the special subunit of PS II, called an “oxygen evolving subunit” containing four manganese atoms. An electron from water is passed to a hydrophobic electron carrier, plastoquinone (PQ) in PS II (electron flow is indicated by thick solid arrows in Fig. 10.1). The PQ that received two electrons gets two protons (H^+) from the water near the membrane surface (the reduced PQ is shown as PQH_2) and hands two electrons over cytochrome (Cyt) b_6f complex. In the latter reaction, PQ releases the protons into the opposite side of the membrane at the same time (uptake and release of protons are indicated by thin gray lines). Cyt b_6f complex passes the electrons to a water-soluble copper-protein, plastocyanin (PC). PC transfers the electrons from Cyt b_6f complex to another chlorophyll-containing complex, photosystem I (PS I). PS I pumps the electron up to ferredoxin (Fd) using light energy. The electrons are, finally, delivered to $NADP^+$ via ferredoxin- $NADP^+$ reductase (FNR), and consequently $NADPH$, a reductive power available to carbon

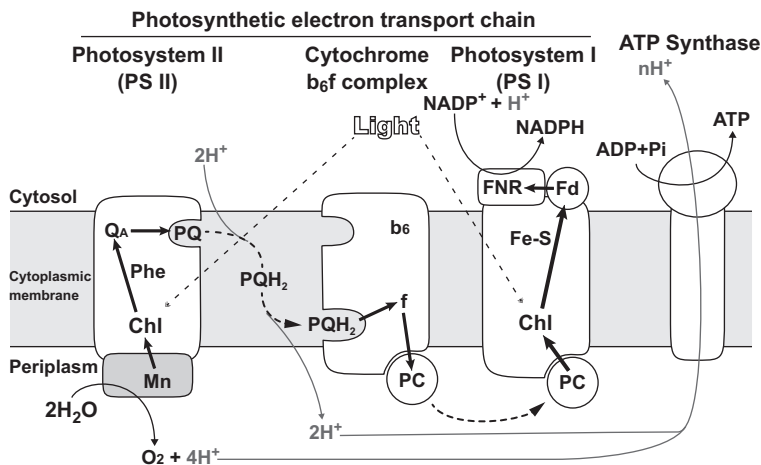


Fig. 10.1 The electron transport system of oxygenic photosynthesis that consists of three membrane-spanning protein complexes, i.e., photosystem II (PS II), cytochrome b_6f , and photosystem I (PS I). Abbrev.: *Mn* manganese in the oxygen evolving subunit, *Chl* chlorophyll molecules, *Phe* pheophytin, *QA* quinone bound to PS II polypeptides, *PQ* plastoquinone, *PQH₂* reduced quinone (quinol), *PC* plastocyanin, *Fe-S* iron-sulfur cluster, *Fd* ferredoxin, *FNR*, ferredoxin-NADP⁺ reductase

fixation, is produced. During electrons are transported among three complexes, the proton gradient is formed across the membrane. The gradient of proton concentration becomes a motive force producing ATP by ATP synthase.

These two photochemical reaction centers, PS I and PS II, share the common features such as possession of the same photopigment (Chl) but are obviously different from each other in respect to their components for transporting an electron and in redox (oxidation-reduction) potential. PS II contains pheophytins and quinones, which are electron carriers with relatively moderate redox potentials. On the other hand, PS I includes iron-sulfur (Fe-S) clusters as electron carriers and can directly reduce NADP⁺ via Fd because PSI has low-potential electron carriers with high reducing power.

10.3 Photosynthetic Reaction Centers in Anoxygenic Photosynthesis

As mentioned above, the electron transport chain of oxygenic photosynthesis has two photosystems (PS I and PS II) to pump up an electron twice with light energy. The double pump-up pathway of electrons found in cyanobacteria (and also in chloroplasts of plants and algae) is called “Z-scheme” based on its shape in the redox diagram (Fig. 10.2). However, anoxygenic photosynthesis, photosynthesis without oxygen production, contains only one of the two photosystems and is not the Z-scheme. In anoxygenic photosynthetic system, electron is transferred cyclic around the photo-induced electron pump. The cyclic electron transported is used to transport

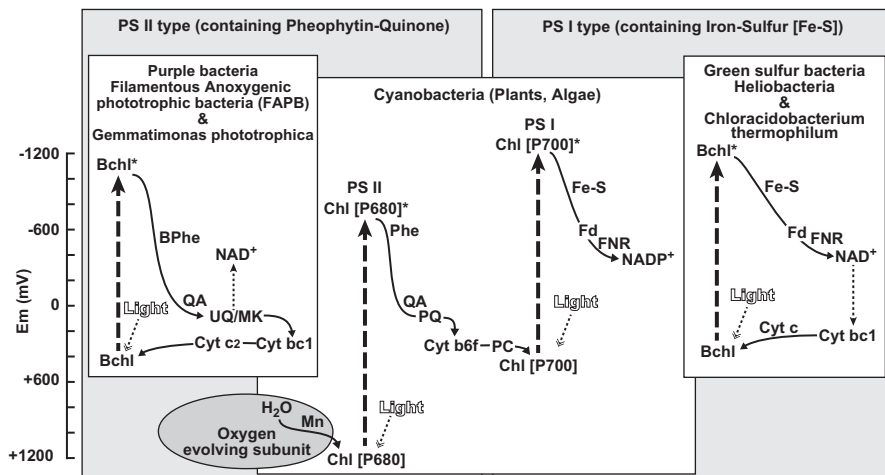


Fig. 10.2 Electron flow systems in photosynthetic organisms. The vertical axis indicates redox (oxidation-reduction) potential (mV). An electron normally flows downward (from negative to positive potential), and only light can pump it up toward high potential

H⁺ across the membrane, which is then used to produce ATP. The system is called cyclic electron transport. Anoxygenic photosynthetic bacteria having one photosystem are widely distributed over the following six bacterial phyla, although oxygenic photosynthesis can be seen only in the phylum *Cyanobacteria* (Castenholz 2015): (1) purple bacteria belonging to the phylum *Proteobacteria* (Madigan and Jung 2009); (2) green sulfur bacteria in the phylum *Chlorobi* (Overmann 2015); (3) filamentous anoxygenic phototrophic bacteria (FAPB) in phylum *Chloroflexi* (Hanada 2014); (4) heliobacteria in the phylum *Firmicutes* (Madigan 2015); (5) *Chloracidobacterium thermophilum* in the phylum *Acidobacteria* (Tank and Bryant 2015); and (6) *Gemmatimonas phototrophica* in the phylum *Gemmatimonadetes* (Zeng et al. 2015).

The PS I-type photosystems are found in green sulfur bacteria and heliobacteria (Fig. 10.2). Fe-S centers are included as electron transport elements in their photosystems; they are able to directly reduce NAD⁺, which is then used to reduce and fix carbon dioxide. In addition to these photosynthetic groups, the similar photosystem is also seen in *Chloracidobacterium thermophilum*. On the other hand, the photosystems in purple bacteria, FAPB, and *Gemmatimonas phototrophica* are closely related to PS II because they all include pheophytins and quinones as electron carriers in their photosystems (Fig. 10.2). However, there is an obvious difference between their photosystems and cyanobacterial PS II. The oxygen evolving subunit including manganese (Mn) atoms that is typical in PS II is absent from the anoxygenic photosystems. Therefore, no anoxygenic photosynthetic bacteria can oxidize water because they lack the oxygen evolving subunit. Instead of water, they use various reductive compounds as electron donors for their photosynthesis, e.g., hydrogen sulfide, thiosulfate, hydrogen, reduced iron, and a wide variety of organic compounds. The phylogenetic positions and features of its photosystem in all oxygenic and anoxygenic photosynthetic bacteria are summarized in Table 10.1.

Table 10.1 The phylogeny of all oxygenic and anoxygenic photosynthetic bacteria and their photosystems

Photosynthetic group	<i>Cyanobacteria</i>	Purple bacteria	Filamentous anoxygenic phototrophic bacteria	Green sulfur bacteria	Heliobacteria	<i>Chloracidobacterium thermophilum</i>	<i>Gemmatimonas phototrophica</i>
Phylogeny (Phyla)	<i>Cyanobacteria</i>	<i>Proteobacteria</i>	<i>Chloroflexi</i>	<i>Chlorobi</i>	<i>Firmicutes</i>	<i>Acidobacteria</i>	<i>Gemmatimonadetes</i>
Phototrophy	Oxygenic	Anoxygenic	Anoxygenic	Anoxygenic	Anoxygenic	Anoxygenic	Anoxygenic
Photosystem(s)	PS I + PS II	PS II	PS II	PS I	PS I	PS I	PS II
Subunit structure of core complex	Heterodimer + heterodimer	Heterodimer	Heterodimer	Homodimer	Homodimer	Homodimer	Heterodimer

10.4 Origin of Oxygenic Photosynthesis

Analysis of mass-independent fractionation of sulfur demonstrated that the rise of atmospheric oxygen occurred approximately 2.1 Gyrs ago (Farquhar et al. 2000). Prior to the oxidation of atmosphere, enormous amounts of reducing compounds dissolved in the ancient ocean were oxidized. Ferrous iron was one of the main reduced constituents and was gradually oxidized and accumulated to the bottom of the ocean along with the increase of oxygen. Oxidized iron was deposited and formed banded iron formation (BIF), and the voluminous BIFs globally developed in the period between 2.5 and 2.3 Gya (Bekker et al. 2010).

The ringleader of the global oxidation was a remote ancestor of cyanobacteria that acquired an oxygenic photosynthetic ability. The production of oxygen was a biological process connected with carbon fixation, and carbon dioxide in the ancient atmosphere was vigorously consumed by the cyanobacterial ancestor to make organic matter. The Huronian glaciation, a global glaciation that extended from 2.4 to 2.1 Gya, happened as a result of the excessive consumption of atmospheric carbon dioxide showing a potent greenhouse effect (Tang and Chen 2013; see Chap. 17 for detail).

The observed fact that the massive BIFs began 2.5 Gya strongly suggests that a cyanobacterial ancestor first invented oxygenic photosynthesis around or shortly before that time. Compared with the history of bacteria (at least 3.8 Gya), the appearance of oxygenic photosynthesis can be said that it is relatively a new event in the long evolutionary process of bacteria.

Anoxygenic photosynthesis has already emerged preceding the emergence of oxygenic photosynthesis (Xiong and Bauer 2002). As mentioned in the previous section, anoxygenic photosynthetic bacteria have developed two different types of photochemical reaction centers, i.e., the pheophytin-quinone-type (PS II-type) and iron-sulfur-type (PS I-type) photosystems, and steadily improved their efficiency and stability under the ancient anoxic environments. PS II-type photosystem is relatively oxidative cyclic electron transport system including pheophytins and quinones. PS I-type photosystem, containing iron-sulfur clusters, was able to directly reduce NAD^+ , which was well adapted to the reductive conditions of the ancient Earth. However, no anoxygenic photosynthetic bacteria in the ancient time could conduct oxygenic photosynthesis until the appearance of a cyanobacterial ancestor.

The most important difference between oxygenic cyanobacteria and anoxygenic photosynthetic bacteria is whether it possesses both two types of photosystems or only either one of the photosystems. In other words, coexisting two different photosystems in a cell are essential for acquisition of oxygenic photosynthesis. The coexistence of photosystems occurred in a cyanobacterial ancestor around 2.5 Gya for the first time, which triggered the development of oxygenic photosynthesis (Blankenship and Hartman 1998). In a cell of the ancestral cyanobacterium that first possessed two different photosystems and allowed their coexistence, the PS I-type photosystem must have been more advantageous for getting energy and fixing carbon dioxide than the PS II-type photosystem, because the former was more adapted to the reductive conditions. Therefore, the PS I-type photosystem would play a main

role in its phototrophic growth. Accordingly, another photosynthetic system, PS II-type photosystem, could be free from the natural selection pressure since it was no longer essential for survival of the ancestral cyanobacterium. This type of photosystem, thus, became easy to change its structure or gene information released from selective pressure and obtained the oxygen evolving subunit for oxygenic photosynthesis at the end (Blankenship and Hartman 1998). Though the evolutionary process for acquisition of ability to use water as an electron donor has not been clear yet, several hypotheses were proposed: Blankenship and Hartman (1998) argued that hydrogen peroxide was the candidate of a transitional electron donor to the PS II-type photosystem preceding the oxygen evolving component; Dismukes et al. (2001) proposed that bicarbonate was a more efficient alternative donor than water for a pre-oxygenic phototroph.

Although the origin of oxygenic photosynthesis is still controversial, it is quite likely that anoxygenic photosynthesis has emerged prior to oxygenic photosynthesis. At least, it is a hint that anoxygenic photosynthetic bacteria possess only one photosystem (a PS I- or PS II-type photosystem), whereas oxygenic phototrophs have both of photosystems.

10.5 Concepts to Interpret Evolution of Photosynthesis

Coexistence of two different photosystems is no doubt indispensable to accomplish oxygenic photosynthesis. Although it is still controversial how two photosystems coexisted in the single cell, two evolutionary models, i.e., selective-loss model (Mulikidjanian et al. 2006; Vermaas 1994) and fusion model (Blankenship 1992; Xiong and Bauer 2002), have been proposed in order to interpret the emergence of photosynthetic systems. In selective-loss model, the coexistence of PS I type and PS II type occurred in the early evolutionary stage via gene duplication, and then an oxygenic phototroph arose from this ancestor, and two types of anoxygenic phototrophs emerged as a result of selectively losing either one of photosystems. While this model has an advantage for explaining the emergence of oxygenic and anoxygenic photosynthesis only with molecular evolution and natural selection, it gives no clear answer to the reason why no anoxygenic phototroph having two different photosystems has been discovered. On the other hand, fusion model is a hypothesis that ancestral PS I-type and PS II-type anoxygenic photosystems are fused to be present in a cell. In this model, the “fusion” is assumed to be a rare phenomenon, which agrees with the fact that oxygenic photosynthesis is limited in a sole phylum, *Cyanobacteria*. According to the hypothesis, two anoxygenic photosystems emerged and could evolve themselves on a long-term basis prior to the origin of oxygenic photosynthesis, supporting that anoxygenic phototrophs have physiological and phylogenetic diversity.

It should be emphasized that the term “fusion” in this model does not mean “fusion of different cells” but mixing two photosystems in a single cell, which is possibly caused by lateral gene transfer (LGT). Actually, some metabolic function, like antibiotics resistance, can be transferred between phylogenetically distant spe-

cies by LGT. Sulfate-reducing ability has been transferred several times not only among distant bacterial phyla but also between the domains *Bacteria* and *Archaea* (Friedrich 2002; Klein et al. 2001). Genes coding for the PS II-type photosystem were transferred within the *Proteobacteria* (Nagashima et al. 1997). Also, the genome analysis of *Chloracidobacterium thermophilum* in the phylum *Acidobacteria* demonstrated that the deduced amino acid sequence of its PS I-type photosystem was closely related to those in green sulfur bacteria belonging to the phylum *Chlorobi* (Bryant et al. 2007), suggesting that *C. thermophilum* received the photosystem from a certain green sulfur bacterium recently. Moreover, it was revealed that the whole photosynthetic gene island with a length of approx. 42.3 kbp (containing all genes related to the photochemical reaction center and biosynthesis of bacteriochlorophyll *a* and carotenoids) was transported at once from purple bacterium belonging to the phylum *Proteobacteria* to *Gemmatimonas phototrophica* in the phylum *Gemmatimonadetes* (Zeng et al. 2014).

PS I-type and PS II-type photosystems are widely distributed among anoxygenic photosynthetic bacteria, and each does not show the monophyletic development, which will be discussed later. Several LGT that occurred within anoxygenic photosynthetic bacteria possibly caused the tangled evolutionary processes of these photosystems.

10.6 Hypothetical Evolutionary Pathway of Photosynthesis Based on Lateral Gene Transfer

Figure 10.3 shows phylogenetic relationship among bacterial lineages containing photosynthetic bacteria based on sequence information of more than 400 broadly conserved proteins (Segata et al. 2013). Photosynthetic bacteria can be found in seven phyla (in Fig. 10.3, lineage names containing photosynthetic species are shown in bold). Though the phyla *Proteobacteria* and *Acidobacteria* are closely related to each other, they are not related to the other phyla. The similar close relationships are seen between the phyla *Chlorobi* and *Gemmatimonadetes* and also between the phyla *Cyanobacteria* and *Chloroflexi*. The phylum *Firmicutes* is the most deeply branched lineage among all phototrophic phyla in this tree.

Oxygenic photosynthesis is seen in the class *Oxyphotobacteria* of the *Cyanobacteria*, and another class in this phylum, *Melainabacteria*, which is a candidate taxon that was inferred from the metagenome analysis, has no photosynthetic genes (Di Rienzi et al. 2013). Anoxygenic photosynthetic bacteria can be found in six phyla. PS I-type phototrophs (the presence of PS I is indicated as a thick broken line in the phylogenetic tree of Fig. 10.3) were found in three phyla: the largest PS I-type phototrophic group designated as green sulfur bacteria belonged to the class *Chlorobea* in the phylum *Chlorobi* (however, the related class *Ignavibacteria* is a non-photosynthetic taxon); *Heliobacteria*, which are included in the phylum *Firmicutes* with a large number of non-photosynthetic bacteria; and *Chloracidobacterium thermophilum*, which is first discovered and the only phototroph in the phylum *Acidobacteria*. On the other hand, PS II-type phototrophs (the

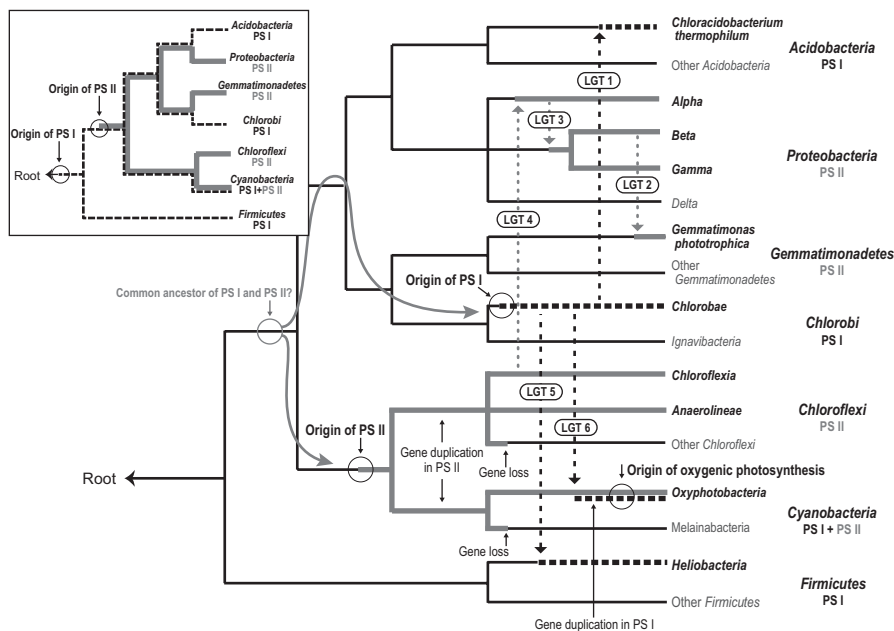


Fig. 10.3 The phylogenetic relationship among bacterial lineages containing phototrophs based on sequence information of more than 400 broadly conserved proteins (Segata et al. 2013). Possession of PS I and PS II is indicated as a “thick broken line” and “thick gray line,” respectively. Transportation of each photosystem by lateral gene transfer (LGT) is shown as a dashed arrow. Inset: hypothetical evolutionary process of photosystems according to the conventional analysis method. This tree was also constructed based on sequences of more than 400 broadly conserved proteins similar to the main phylogenetic tree and was simplified by removing the non-phototrophic lineages

presence of PS II is indicated as a thick gray line) can be found in the following phyla: in the phylum *Proteobacteria*, three classes (*Alpha-*, *Beta-*, and *Gammaproteobacteria*) contain purple photosynthetic bacteria (while not in *Deltaproteobacteria*); other PS II-type phototrophic groups called FAPB exist in the classes *Chloroflexia* and *Anaerolineae* within the phylum *Chloroflexi*; and *Gemmatimonas phototrophica* is the sole phototroph in the phylum *Gemmatimonadetes*.

The only one photosynthetic lineage that has two types of photosystems is cyanobacteria (to be precise, the class *Oxyphotobacteria*); all photosynthetic bacteria other than cyanobacteria possess a single photosystem, i.e., either PS I- or PS II-type photosystem. Although anoxygenic photosynthesis is complicatedly distributed among bacterial phyla, oxygenic photosynthesis is limited to the lineage of *Oxyphotobacteria*. Assuming that coexistence of two photosystems is essential for the emergence of oxygenic photosynthesis and scarcely happens, two different photosystems must have met on a branch of the *Oxyphotobacteria* lineage for the first time. If the phenomenon occurs for many times throughout long bacterial history, oxygenic photosynthesis no doubt must have emerged on other branches in the phy-

logenetic tree of bacteria. However, no such organism has been found yet. Therefore, it is likely that the coexistence of two photosystems occurred in the long evolutionary history firstly and only once in *Oxyphotobacteria*.

The origin and evolutionary process of each type of photosystems are considered in Fig. 10.3. According to the conventional methods in phylogenetic analysis (Fig. 10.3 inset), the origin of each photosystem is placed on the intersection node of the all lineage that possesses it. Thus, the origin of PS I-type photosystem is on the root of the tree and that of PS II-type photosystem is on the deepest branch. Based on the result, it can be interpreted that the origin of PS I-type photosystem is older than that of PS II type. The phylogenetic interpretation also shows that the coexistence of the two photosystems occurred several times, and which conflicts with the primary assumption that the coexistence is assumed a quite rare phenomenon.

Lateral gene transfer (LGT) is a solution to explain evolutionary processes of two photosystems without any contradiction. As mentioned in the previous section, genes encoding the photosystems are transported occasionally from a phototroph to the other. *C. thermophilum* belonging to the phylum *Acidobacteria* received its PS I-type photosystem by LGT from a green sulfur bacterium in the class *Chlorocheae* indicated as “LGT 1” in Fig. 10.3 (Bryant et al. 2007). Likewise, *G. phototrophica* in the phylum *Gemmatimonadetes* accepted the PS II-type photosystem from a purple bacterium in the class *Betaproteobacteria* shown as “LGT 2” in Fig. 10.3 (Zeng et al. 2014).

A core complex in PS II-type photosystem is a heterodimer resulted from gene duplication (Blankenship 1992). A recent phylogenetic analysis based on deduced amino acid sequence of the core complex genes (Cardona 2015) revealed that (1) these sequences bore resemblance to each other, (2) the gene duplication occurred twice in two separate lineages, and (3) the duplicated genes were transferred by LGT between the phyla *Proteobacteria* and *Chloroflexi* after the gene duplication. In addition, Nagashima et al. (1997) demonstrated that the genes encoding for PS II-type photosystems in *Beta*- and *Gammaproteobacteria* were originated in *Alphaproteobacteria* (“LGT 3” in Fig. 10.3). This line of evidence strongly suggests that the origin of proteobacterial photosystems must be located in a branch of *Alphaproteobacteria* but not at a position deeper than an intersection node of the phylum *Proteobacteria*. Otherwise, purple bacteria in *Beta*- and *Gammaproteobacteria* could once lost their own photosystems and then obtained the similar one from *Alphaproteobacteria*, which is parsimoniously unlikely. On the other hand, no evidence of LGT of PS II types has been found in the phylum *Chloroflexi*. Therefore, conclusions inferred from these phylogenetic findings suggest the model as follows: (1) the PS II-type photosystem appeared in the common ancestor of the phyla *Cyanobacteria* and *Chloroflexi* (“Origin of PS II” in Fig. 10.3) and (2) the proteobacterial PS II-type photosystems were received from the phylum *Chloroflexi* (“LGT 4” in Fig. 10.3). Within the phylum *Chloroflexi*, the photosynthetic genes were lost in lineages other than the Class *Chloroflexia* and *Anaerolineae*. Similar gene loss has occurred on the branch of the Class *Melainobacteria* in the phylum *Cyanobacteria*.

It is also unlikely that PS I type appeared in the common ancestor of the phyla *Chlorobi*, *Firmicutes*, and *Cyanobacteria* as shown in inset of Fig. 10.3. The photosynthetic group, *Heliobacteria*, consists of only three species, while its phylum *Firmicutes* is a big bacterial taxon containing approx. 2500 species. If we assume that an ancestor of *Firmicutes* species was a phototroph, we must accept that almost all species (approx. 99% of the total species) other than *Heliobacteria* have lost their phototrophy. It is more likely that an ancestor of *Heliobacteria* received the photosystem from the other lineage, a *Chlorobi* species, by LGT (“LGT 5” in Fig. 10.3). From this assumption, it can be inferred that PS I-type photosystem originated in a common ancestor of *Chloroidea* (“Origin of PS I” in Fig. 10.3), and thus the origin of PS I-type photosystem is later than that of PS II type.

According to the hypothesis shown in Fig. 10.3, a cyanobacterial ancestor was a PS II-type phototroph, and the ancestor acquired oxygenic photosynthetic ability after getting PS I-type photosystem from an ancient green sulfur bacterium in the phylum *Chlorobi* (“LGT 6 in Fig. 10.3). Whereas the present cyanobacterial PS I has heterodimeric core complex, the PS I-type photosystem of green sulfur bacteria still retains a homodimeric core complex. These suggest the duplication of a PS I encoding gene occurred in the cyanobacterial lineage after its acquisition by LGT.

The evolutionary process of photosystems is summarized in Fig. 10.4. On the ancient Earth, PS II-type photosystem (homodimer) has been first emerged in an ancestor of the phyla *Cyanobacteria* and *Chloroflexi*, and the primitive PS II was improved and changed into a heterodimeric form both in the lineages independently (these polypeptides were designated L and M in the *Chloroflexi* lineage, D1 and D2 in the phylum *Cyanobacteria*, respectively, as shown “L/M” and “D1/D2” in Fig. 10.4). After the emergence of PS II type, PS I-type photosystem appeared in the *Chlorobi* lineage. This new PS I-type photosystem contained Fe-S clusters and was well adapted to anoxic environments in the ancient Earth. Meanwhile, heterodimeric PS II-type photosystem was transferred from the *Chloroflexi* lineage to the *Proteobacteria* lineage.

In about 2.5 Gyrs ago, the oxygen evolving subunit was at last invented through trial and error, and oxygenic photosynthesis that is able to use water as an electron donor was established. After the establishment, oxygen tension was gradually increased, and the environments became aerobic. PS II-type photosystem containing quinone as an electron carrier was able to tolerate oxygen, and many PS II-type phototrophs acquired oxygen respiration ability in addition to anoxygenic phototrophy (Blankenship 1992; Xiong and Bauer 2002). *G. phototrophic* also received the photosystem from a purple bacterium in the *Proteobacteria* lineage similar timing.

On the other hand, most PS I-type phototrophs had no choice but to run away from oxygen due to high susceptibility of their Fe-S containing PS I to oxygen. One exception was *C. thermophilum* in the phylum *Acidobacteria*. The system was oxygen resistant from the beginning because the photosystem was acquired after the increase of oxygen tension.

It is still controversial whether two types of the photosystems have a common ancestor or not. Two photosystems share common features in structural and pigmentary respects (Blankenship 1992; Xiong and Bauer 2002), suggesting the presence of a common ancestor between them. However, the two types of photosystems

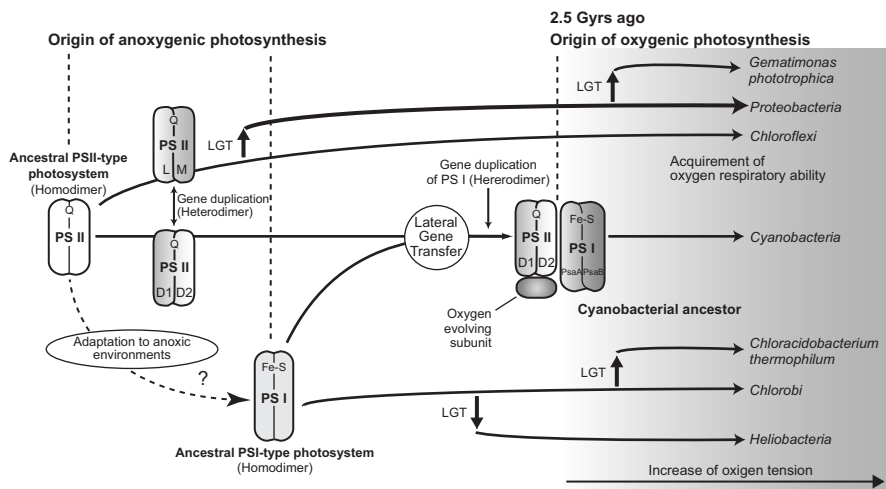


Fig. 10.4 A hypothetical evolutionary process of photosystems proposed in this chapter. It is still open to question whether an ancestral PS I-type photosystem evolved from ancestral PS II-type photosystem or appeared independently. Abbrev.: *LGT* lateral gene transfer, *Q* quinone, *Fe-S* iron-sulfur cluster

may have appeared independently, because their amino acid sequences of peptides that form the photosystems show quite low similarities (Sadekar et al. 2006).

Assuming that coexistence of two different photosystems is rare phenomenon and emergence of oxygenic photosynthesis occurred only once in the Earth's history, this hypothetical model based on LGT can explain enigmatic evolution of photosystems without contradiction. As long as no oxygenic photosynthetic prokaryote other than cyanobacteria would be found, the hypothesis can be the good starting point.

10.7 Role of Anoxygenic Photosynthesis in Ancient Earth

Anoxygenic photosynthesis no doubt played an important role in anoxic environments of the ancient Earth. They were able to oxidize a variety of compounds by light even under anoxic conditions. A number of anoxygenic phototrophs can use reductive sulfur compounds, e.g., hydrogen sulfide and elemental sulfur, as an electron donor, and convert them into sulfate. Sulfate produced by anoxygenic phototrophs was utilized by sulfate-reducing bacteria as an electron acceptor. Since reduced sulfur compounds are hardly oxidized in the anoxic environments, successive supply of sulfate to the ancient sulfate reducers must have been relied on anoxygenic photosynthesis.

Anoxygenic phototrophs can also oxidize ferrous iron (Fe^{2+}). Ehrenreich and Widdel (1994) discovered anaerobic oxidation of ferrous iron by purple bacteria that belonged to the phylum *Proteobacteria*. The similar ability was found in green

sulfur bacterium in the phylum *Chlorobi* (Heising et al. 1999). These anoxygenic phototrophs are able to oxidize ferrous iron and to produce a large amount of ferric deposits, and which probably caused the Archean banded Iron formation that occurred prior to emergence of oxygenic photosynthesis (Xiong and Bauer 2002). The ferric compounds were also useful for organisms living by anaerobic respiration of iron.

In anoxic environments, anoxygenic photosynthesis was the sole mechanism that was able to oxidize reduced compounds and certainly supported anaerobic respirators in the ancient microecosystem consisting of sulfate and iron reducers. Anoxygenic photosynthesis was indispensable to the ancient environments as an oxidizing system of various reduced compounds with light energy.

10.8 Conclusions

Oxygenic photosynthesis has emerged from anoxygenic photosynthesis in the long history of Earth. Prior to the emergence of oxygenic photosynthesis, two types of photosystems, i.e., PS I-type and PS II-type photosystems, have emerged and evolved in anoxygenic photosynthetic bacteria in the anoxic environments. Coexistence of the different photosystems in a single cell that occurred by lateral gene transfer (LGT) triggered the emergence of oxygenic photosynthesis. A cyanobacterial ancestor that originally possessed PS II-type photosystem received a gene for PS I-type photosystem at least 2.5 Gyrs ago. Such LGT frequently occurred in the evolutionary process of anoxygenic photosynthesis, which made a variety of anoxygenic phototrophs with a wide phylogenetic diversity on Earth.

The Goldilocks zone means the place not only at “the just right temperature” but also under the adequate irradiation of light. Therefore, there is a high possibility that anoxygenic photosynthesis emerged on every planet in the Goldilocks zone because it is the sole mechanism to oxidize reductive compounds under the anaerobic conditions and can maintain the ecosystem consisting of anaerobes by continuously providing oxidative compounds to anaerobic respirators. In these Goldilocks planet other than Earth, however, it is not clear whether oxygenic photosynthesis could evolve from anoxygenic photosynthesis. Although oxygenic photosynthesis has a great advantage using water that presents everywhere as an electron donor, its acquisition requires coexistence of two different photosystems and development of a complete oxygen evolving subunit. The findings in the evolutionary process of photosynthesis inferred that the coexistence of photosystems is a quite rare opportunity and it is difficult to obtain the oxygen evolving subunit, though it may be possible from the billions of evolutionally trials similar to what happened on Earth.

In near future, the development of astronomical research will enable to know the atmospheric states of several Goldilocks planets with cutting-edge techniques for remote sensing (see Chap. 29). If chlorophylls or the related photosynthetic pigments were to be detected in the planets with spectroscopic sensing, it will suggest that anoxygenic photosynthesis has been emerged. If the atmosphere containing

oxygen were to be found, we will know that the oxygenic photosynthesis has evolved on the planet. If the atmosphere of these all planets include oxygen without exception, it will suggest that emergence of oxygenic photosynthesis on Earth does not occurred by chance but is an inevitable evolutionary process.

References

- Bekker A, Slack JF, Planavsky N et al (2010) Iron formation: the sedimentary product of a complex interplay among mantle, tectonic, oceanic, and biospheric processes. *Econ Geol* 105:467–508
- Blankenship RE (1992) Origin and early evolution of photosynthesis. *Photosynth Res* 33:91–111
- Blankenship RE, Hartman H (1998) The origin and evolution of oxygenic photosynthesis. *Trends Biochem Sci* 6:4–6
- Bryant DA, Costas AMG, Maresca JA et al (2007) *Candidatus Chloracidobacterium thermophilum*: an aerobic phototrophic Acidobacterium. *Science* 317:523–526
- Cardona T (2015) A fresh look at the evolution and diversification of photochemical reaction centers. *Photosynth Res* 126:111–134
- Castenholz RW (2015) General characteristics of the cyanobacteria. In: Whitman WB (ed) *Bergey's manual of systematic bacteriology*. Wiley, New York, pp 1–23
- Des Marais DJ (2000) When did photosynthesis emerge on earth? *Science* 289:1703–1705
- Di Rienzi SC, Sharon I, Wrighton KC et al (2013) The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to cyanobacteria. *elife* 2013:1–25
- Dismukes GC, Klimov VV, Baranov SV et al (2001) The origin of atmospheric oxygen on earth: the innovation of oxygenic photosynthesis. *Proc Natl Acad Sci U S A* 98:2170–2175
- Ehrenreich A, Widdel F (1994) Anaerobic oxidation of ferrous iron by purple bacteria, a new-type of phototrophic metabolism. *Appl Environ Microbiol* 60:4517–4526
- Farquhar J, Bao H, Thieme MH (2000) Atmospheric influence of earth's earliest sulfur cycle. *Science* 189:756–759
- Friedrich M (2002) Phylogenetic analysis reveals multiple lateral transfers of adenosine-5-phosphosulfate reductase genes among sulfate-reducing microorganisms. *J Bacteriol* 184:278–289
- Hanada S (2014) The phylum Chloroflexi, the family *Chloroflexaceae*, and the related phototrophic families *Oscillochloridaceae* and *Roseiflexaceae*. In: Dworkin M, Falkow S, Rosenberg E et al (eds) *Prokaryotes*, Proteobacteria delta Epsil. subclasses. *Deep. rooting Bact*, vol 7. Springer, Berlin, pp 515–532
- Heising S, Richter L, Ludwig W, Schink B (1999) *Chlorobium ferrooxidans* sp. nov., a phototrophic green sulfur bacterium that oxidizes ferrous iron in coculture with a “*Geospirillum*” sp. strain. *Arch Microbiol* 172:116–124
- Klein M, Friedrich M, Roger AJ et al (2001) Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes. *J Bacteriol* 183:6028–6035
- Lyons TW, Reinhard CT, Planavsky NJ (2014) The rise of oxygen in earth's early ocean and atmosphere. *Nature* 506:307–315
- Madigan MT (2015) *Heliobacterium*. In: Whitman WB (ed) *Bergey's Manual of Systematic Bacteriology*. Wiley, New York, pp 1–4
- Madigan M, Jung DO (2009) An overview of purple bacteria: systematics, physiology, and habitats. In: Hunter CN, Daldal F, Thurnauer MC, Beatty JT (eds) *The purple phototrophic bacteria*, *Advances in photosynthesis and respiration*, vol 28. Springer, Dordrecht, pp 1–15
- Mulkidjanian AY, Koonin EV, Makarova KS et al (2006) The cyanobacterial genome core and the origin of photosynthesis. *Proc Natl Acad Sci* 103:13126–13131
- Nagashima KVP, Hiraishi A, Shimada K, Matsuura K (1997) Horizontal transfer of genes coding for the photosynthetic reaction centers of purple bacteria. *J Mol Evol* 45:131–136

- Overmann J (2015) Green sulfur bacteria. In: Whitman WB (ed) *Bergey's manual of systematic bacteriology*. Wiley, New York, pp 1–8
- Sadekar S, Raymond J, Blankenship RE (2006) Conservation of distantly related membrane proteins: photosynthetic reaction centers share a common structural core. *Mol Biol Evol* 23:2001–2007
- Segata N, Börnigen D, Morgan XC, Huttenhower C (2013) PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* 4:2304
- Tang H, Chen Y (2013) Global glaciations and atmospheric change at ca. 2.3 Ga. *Geosci Front* 4:583–596
- Tank M, Bryant DA (2015) *Chloracidobacterium thermophilum* gen. nov., sp. nov.: an anoxygenic microaerophilic chlorophotoheterotrophic acidobacterium. *Int J Syst Evol Microbiol* 65:1426–1430
- Vermaas WF (1994) Evolution of heliobacteria: implications for photosynthetic reaction center complexes. *Photosynth Res* 41:285–294
- Xiong J, Bauer CE (2002) Complex evolution of photosynthesis. *Annu Rev Plant Biol* 53:503–521
- Zeng Y, Feng F, Medova H et al (2014) Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. *Proc Natl Acad Sci* 111:7795–7800
- Zeng Y, Selyanin V, Lukes M et al (2015) Characterization of the microaerophilic, bacteriochlorophyll a-containing bacterium *Gemmatimonas phototrophica* sp. Nov., and emended descriptions of the genus *Gemmatimonas* and *Gemmatimonas aurantiaca*. *Int J Syst Evol Microbiol* 65:2410–2419

Chapter 11

Cosmolinguistics: Necessary Components for the Emergence of a Language-Like Communication System in a Habitable Planet



Kazuo Okanoya

Abstract The emergence of human language is one of the biggest wonders in the universe. In this chapter, I define “a language-like communication system” and examine the components necessary for the emergence of such a system, not only on Earth but in any habitable planet. Human language is a unique system among animal communication. Language is a system of transmitting an infinite variety of meanings by combining a finite number of tokens based on a set of rules. Language is not only used in communication but also in thinking. Thus, language is a system that enables compositional semantics. I propose that at least three components are necessary for the emergence of a language-like system: segmentation of context and behavior, the association between them, and the honesty of the emitted signals. When a signal conveys sufficient information regarding the behavioral state of the sender, that signal is defined as “honest,” meaning that its production incurs physiological, temporal, and social costs. I explain each of these conditions and discuss the possibility of “language as it could be” on other planets. I also extend my argument to the future of linguistic communication.

Keywords Segmentation · Association · Signal honesty · Ritualization · Communication · Extraterrestrials

11.1 Introduction

Communication with extraterrestrial existence is a favorite topic of science fiction novels. The novel *Solaris*, written by Stanislaw Lem in 1961, describes the entire planet Solaris as an intelligent being. The planet Solaris tries to study the very human researchers who are, in turn, trying to study it (Lem 1961). Solaris seems to

K. Okanoya (✉)
The University of Tokyo, Tokyo, Japan
e-mail: cokanoya@mail.ecc.u-tokyo.ac.jp

be sentient and reactive to human investigation; however, the attempt to “communicate” with the planet fails. This is partially due to Solaris’s reflective nature: the aim of communication for Solaris is to mimic the inside of human mind, while for humans, communication is believed to be mutually beneficial. More recently, Ted Chiang wrote the novel *Story of Your Life*, in which heptapod extraterrestrials visit and try to communicate with humans in their specific “language” (Chiang 1998). Their language has a holistic nature that allows it to transcend the time dimension. By studying their language, the linguist acquires a unique perception of time; he transcends the time dimension and sees the past and future simultaneously. These novels present possible structures of extraterrestrial language.

Here, I define language as a system of transmitting an infinite variety of meanings by combining a finite number of tokens based on a set of rules. Each token, in turn, has multiple associations with specific meanings. Language is not only used in communication but also in thinking. In fact, in some schools of theoretical linguistics, language is considered to have originated as a tool for thought (Chomsky 2000). Thus, language is a system that enables not only communication but also compositional semantics. How could such a system have evolved on planet Earth, and what conditions would be necessary for such a system to evolve outside of Earth? By asking these questions, I aim to start a new branch of linguistics, namely, “Cosmolinguistics.”

11.2 The Emergence of Language-Like Communicative Signals on Earth

Communication in the context of biology is defined as “the transmission of a signal from one animal to another such that the sender benefits, on average, from the response of the recipient” (Slater 1983). Since this definition does not include an intention on the part of the signaler or a benefit to the receiver, it is useful to avoid anthropomorphic interpretations of animal behavior. Anthropomorphic views include the false notion that communication is mutually beneficial and communication is an indication of self-awareness. Communication can, in fact, evolve without self-awareness and mutual benefit (Bradbury and Vehrencamp 2012). In this section, I briefly propose a set of hypotheses to account for the emergence of language on planet Earth. Here, I limit myself to the discussion of acoustic communication in vertebrates, because the principal medium for language remains speech communication. I am aware that this is a specific condition on Earth in which most animals require respiration for metabolism.

11.2.1 *Ritualization of Respiratory Movements*

Communicative acoustic signals have always started as a secondary trait in vertebrate animal behavior (Fitch and Hauser 2003). Acoustic signals often originate from respiratory actions because respiratory organs function as air passages.

Because respiration is an action that is absolutely vital for animal survival, the secondary use of respiratory energy for vocal production has a low physiological cost (Oberweger and Goller 2001). The respiratory tract is a pipe connecting the bilateral lungs and mouth opening. Because the respiratory tract extends into the body, physiological conditions affect its acoustical characteristics. Coughing is associated with infection and inflammation of the respiratory tract. Strong exhalation produces noise associated with the length of the respiratory tract (Morton 1977). Furthermore, because opening the mouth is preparatory behavior for biting or attacking in predatory animals, the exhalative noise associated with mouth opening could signal attack (Briefer 2012). In this way, respiratory noise is correlated with subsequent behavior by the signaler.

When such signals change the behavior of the receiver so that the change benefits the sender, the signals gain communicative value. For example, the pup isolation calls of rodents comprise short, repeating ultrasonic calls. This acoustic signal has a characteristic of easy localization because, due to the short wavelength of ultrasounds, there are many onset-offset cues with phase information available for the small rodent heads. Upon detecting the isolation call, the mother quickly approaches to retrieve the pup, who is the sender of the call (Ehret 2005). These calls must have originated from the respiratory noises arising from the short and shrunken tracheae of infant animals, whose body temperature has quickly fallen due to isolation from the mother. Calls must then have undergone natural selection for localizability. During this process, the noise that originated due to hypothermia must have become the isolation call (Fig. 11.1).

11.2.2 *Emergence of Songs*

Most land animals emit “calls” specific to behavioral contexts. Calls are monosyllabic, simple vocalizations. In addition to calls, some animals emit trains of various calls, and such vocalizations are often used in mating contexts. Because of the acoustic resemblance to human singing, these vocalizations are sometimes referred to as songs. It has remained an enigma how songs emerged in animals.

In rodents, when pups are out of the nest, they emit isolation calls that induce retrieval responses from the mother. When bird chicks are hungry, they emit food-begging calls to make their parents bring food to them. When human babies need physiological or social care, they emit baby cries. These care-inducing signals are always in the form of repeated calls. This is true in rodents, birds, and humans (Wright and Leonard 2007). While repeated signals may increase the chance of detection, they may also increase the risk of becoming habituated. Infant pygmy marmosets produce repeated vocalizations when seeking care from adult animals, but they do so by combining different calls (Elowson et al. 1998). These call-repeating behaviors in young animals might be a preadaptation of songs in adults. Because these behaviors mimic infantile behavior, a tendency to produce randomly repeated calls may induce a strong reaction in female listeners.

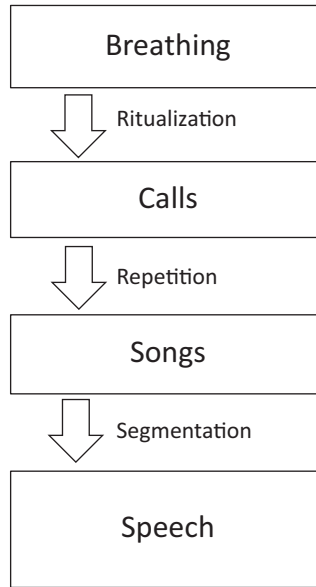


Fig. 11.1 Schematic account of the set of hypotheses accounting for the emergence of language on planet Earth. Acoustic communication began as an expression of emotion associated with breathing. Such signals then became ritualized and the action patterns were fixed as calls. Repetitive calls were used by infant animals to intensify their signal value to mothers or parents. Similar signals were then mimicked by adult animals to relax female listeners in mating context. These signals comprise songs. Most animals sing innate songs, and receivers began to extract honest information about individual vigor from these songs. Songs then became sexually selected traits. In some species, complexity was preferred as a signal of vigor, and songs became a learned trait allowing further complexity. Such complex learned songs were shared in the societies of protohumans. The mutual segmentation of behavioral contexts and song phrases led to the emergence of speech

Supporting evidence for this infantile mimicry hypothesis comes from a neuro-anatomical study in the songbird brain (Liu et al. 2009). Chicks of chipping sparrows produce variable sequences of food-begging calls. When an expression of an immediate early gene (gene activated immediately after neural firing) was examined in the brain of these chicks, the area corresponding to the adult RA (robust nucleus of the arcopallium, homologues to the motor cortex in mammals) showed strong activation. Partial lesions of the same area resulted in a reduction in the variability of food-begging calls. The results indicate that food-begging and adult songs may utilize the same neural resources. This finding supports the hypothesis that food-begging calls may be a preadaptation to songs in birds.

Another line of evidence includes neurophysiological studies with mammalian isolation calls, including human cries. In rats and squirrel monkeys, lesioning the anterior cingulate cortex resulted in changes in the acoustic structures of isolation calls. In human babies, neural activity induced by crying was observed in the same brain area (Newman 2007). In adult mice, lesioning the anterior cingulate cortex resulted in changes in temporal and acoustical structures in courtship songs (Ariaga

unpublished observation). On the other hand, a mutant mouse that lacked neocortical and hippocampal areas sang normal songs, suggesting that only a part of the cortex may be necessary for courtship songs (Hammerschmidt et al. 2015).

Some species of bird, cetacean, and bat, in addition to one primate (only humans) demonstrate the additional faculty of vocal learning (Jarvis 2006). Vocal learning is the ability to acquire a new vocal repertoire through auditory-motor feedback learning. Vocal learning enables song complexity and eventually syllable variety in human speech. When and how vocal learning evolved is not known, but several hypotheses have been proposed, including mother-offspring interaction, sexual selection, domestication, and antipredator defense (Okanoya 2017).

Taken together, the idea that isolation calls and food-begging calls might be precursors to adult mating songs is consistent with the current data on neural mechanisms for vocal productions. Further studies are necessary to relate isolation and food-begging calls with adult mating songs in birds and mammals.

11.2.3 *Emergence of Speech*

Here, my challenge is to place the emergence of human speech in a continuous evolutionary line with the emergence of songs and the evolution of song complexity in nonhuman animals. To demonstrate the continuum of development with other primates, I will first examine song-like behavior in nonhuman primates and then propose a hypothesis related to the emergence of speech out of songs.

Gibbons are one of the five ape groups, of which the other four are humans, chimpanzees, gorillas, and orangutans. Because they are not great apes, gibbons are the most distant of the apes from humans. Gibbons do have song-like vocalizations (Geissmann 2002), but they are not learned, as indicated by cross-fostering studies. Cross-fostering studies involve exchanging babies of two species immediately after they were born and rearing the babies by the species different from their genetic species. Cross-fostering between two species of gibbons showed that gibbon songs are genetically determined and no effect of rearing environment was observed (Merker and Cox 1999). Nevertheless, gibbon songs are relatively diverse (Clarke et al. 2006) and are not only used in a mating context but also in many other social contexts (Inoue et al. 2012). In Muller gibbons, male calls consist of two simple types: a frequency-modulated “wa” call and a constant “o” call. Combinations of these calls and behavioral contexts have been correlated, meaning that gibbons might exchange contextual information via the combination of calls.

The gelada is a species of primate with a rich vocal repertoire. They also make a facial expression, with lip-smacking of 3–8 Hz used as an affiliative signal. On some occasions, their lip-smacking is presented with vocal sounds, making this behavior highly similar to human speech production (Bergman 2013). Other primates including macaques also show lip-smacking, and this behavior might be one of the precursors to human speech (Ghazanfar et al. 2012).

Both of these behaviors, vocal repertoire and lip-smacking, if combined with the bird-like ability of vocal learning, would provide a basis for the emergence of human speech.

11.3 Components for the Emergence of Language-Like Systems

I will now try to extend what might happen on Earth to habitable planets in general. Life evolved on planet Earth under highly specific conditions. Human language is the product of complex and arbitrary historical interactions that occurred only once on Earth. Nevertheless, it is possible to specify the boundary conditions that would lead to the emergence of language-like communication system on potentially habitable planets. I suggest that segmentation, association, and signal honesty are three key components necessary for such emergence.

Although vocal learning is considered a necessary condition for the emergence of human speech (Deacon 1998), I do not consider that condition at this point. This is because, in theory, language-like communication is possible if the agent has at least a binary (1 or 0) signal that is innately prepared. As is evident from digital computer architecture, a binary signal can emulate any degree of complexity. It is true that vocal learning and signal complexity can compress the time required to convey information, but these time constraints could vary depending on the agent's sensory and motor capacity.

11.4 Segmentation

Given a string of behavioral sequence, such as that for song on Earth, if there is a behavior that is sequentially or spatially emitted, it would provide the basis for segmentation and chunking. Segmentation is the process of cutting down longer or larger entities into shorter or smaller pieces. Chunking, on the other hand, is the process that does the opposite: amalgamating pieces to create a longer or larger entity. Segmentation occurs both in auditory and visual domains of the brain in vertebrate animals by means of lateral inhibition, in which the neurons that fired inhibit the activity of neighboring neurons (Meinhardt and Gierer 2000) or statistical learning, in which transition probabilities between two successive stimuli are learned (Saffran et al. 1999). The external environment would usually be more or less continuous, but living agents that move around must be able to segment or categorize the environment in order to reduce the load for sensory information processing. Segmentation is crucial in any agent that moves around a nonuniform environment.

When communication among similar organisms or conspecifics becomes beneficial, then the organisms output stimuli that are perceived by other organisms. These stimuli may be long and continuous on a physical domain but are packaged or chunked into pieces based on the motor constraints of their producing agent. The receiver may also segment the physically continuous stimuli into perceptual chunks based on the sensory constraints. The stimuli then become signals. In this way, the chunking of behavioral units and the segmenting of the perceived unit occur among the communicating agents.

Segmentation and chunking not only occur on stimuli but also in behavioral contexts. A behaving agent should know which behavioral situation is occurring in a given moment. This ability will also reduce the variety of its own behavioral state and make it easier to associate a stimulus with a behavior.

Consider what might have happened on Earth. How might song-like behavior in some primate species be connected with speech in humans? We proposed a conceptual model for this process (Merker and Okanoya 2007), in which each behavioral context is denoted by a particular song in a protohuman society. Consider the hypothesis that prior to language, protohumans developed singing behavior associated with several social contexts. If songs became a learned property, as they are in some species of bird and whale, a syllable phrase may be shared by more than one song. Then, likewise, parts of the behavioral contexts in which a song is sung may also be shared by more than one song (Fig. 11.2).

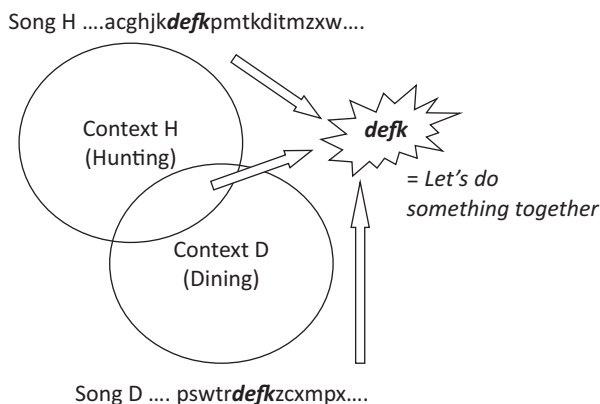


Fig. 11.2 Mutual segmentation of song phrases and behavioral contexts. When two songs share a common phrase and context in which the songs are sung, the song (part) phrase and (part) context are mutually segmented and associated. The very short segmented song phrase then comes to denote the segmented specific context. In the specific example provided here, song H is sung when agents go hunting, and song D is sung when agents go dining. Because one of the contexts common to hunting and dining is “doing something together,” the shared part of the two songs “defk” would likely be associated with the meaning of “doing something together” in the next generation of agents. In this way, holistic songs were gradually segmented into shorter pieces and associated with specific meanings through generations of agents

For example, a song sung when hunting (song H) and a song sung when dining (song D) might have shared the same phrase h&d. Furthermore, song H and song D shared the context of doing something together. After a while, by singing the shared phrase h&d, the singer could have specified the context of “let’s do that together.” By repeating this process, holistic songs might have been decomposed into specific phrases, which may in turn have become proto-words.

I call this the mutual segmentation hypothesis of song phrases and song contexts (Merker and Okanoya 2007; Okanoya and Merker 2007). Once the process of mutual segmentation commenced and segmented short utterances became associated with segmented restricted contexts, rudimentary forms of speech communication could have commenced. Subsequently, non-biological, cultural processes came to regulate the emergence of syntactical structures.

Segmentation and chunking in the signal and behavioral domain are thus essential for the evolution of language-like communication systems.

11.4.1 Association

In the previous section, I automatically assumed this faculty of associating stimuli and behavior. In all animals on Earth, associating given stimuli with given behaviors is an essential capacity for survival. This is shown to exist already in single-cell animals. The simplest form of such an association is habituation, in which repeated exposure to a given stimulus results in a reduction of behavioral response (Castellucci et al. 1970). A more complex form of association is known as Pavlovian conditioning, in which a neutral stimulus gains signal value through association with a key stimulus that can innately induce a certain response (Rescorla 1972). Operant conditioning is a further advanced form of associative learning in which the probability of occurrence of a defined behavior changes through external reward (Skinner 1990). For mutual segmentation of string and context to occur, I am assuming that at least associating a part of a string with a part of a context would be beneficial for the organism. Associative learning should be adaptive in any agents, as it affords the opportunity to predict what will occur next, as well as the selectivity to choose stimuli that result in positive reinforcement, and to avoid stimuli that result in punishment (Fig. 11.3).

11.4.2 Signal Honesty

For the receivers of the signal, it is crucial that the signal reflects the true behavioral state of the sender. If not, the signal loses its value and gradually ceases to function. Behavioral states include emotional, intentional, nutritional, and genetic (Brudzynski 2014; Searcy and Nowicki 2005). When a signal conveys sufficient information regarding the behavioral state of the sender, it is defined as “honest” (Searcy and

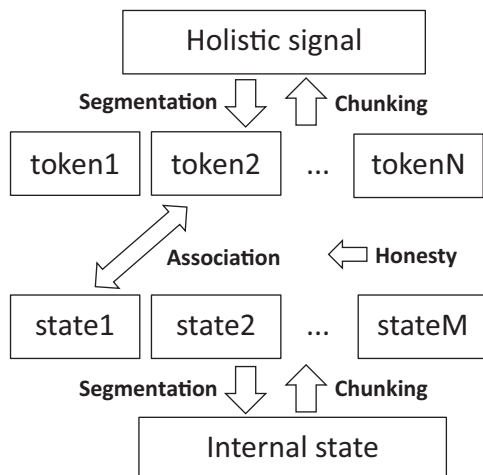


Fig. 11.3 Necessary components to form a language-like communication system. Signals should have a hierarchy, which is achieved by chunking components (token1, token2, ... tokenN) into a holistic signal or segmenting a holistic signal into tokens. Likewise, internal states should also have a hierarchy by chunking and segmentation. Internal states correspond to behavioral contexts interpreted by the agent. Behavioral tokens and internal states are associated via temporal proximity to form token-state pairs. Tokens must be honest signals if they are to guarantee the occurrence of a certain internal state. Thus, the production of each token is costly

Nowicki 2005). An honest signal bears “costs” of producing, such as physiological, temporal, and social costs. For example, birdsong incurs costs in terms of neural resources, metabolism, the risk of being located by a predator, and time costs (e.g., reduced time for foraging or other alternative behaviors). Thus, singing can be an honest signal to indicate the singer’s resourcefulness and fitness. The above considerations on signal honesty should apply in any biological system that evolves not only on Earth but on any habitable planet.

11.5 Cosmolinguistics

I have discussed three components necessary to form a language-like communication system: segmentation, association, and signal honesty. Language is a system of transmitting an infinite variety of meanings by combining a finite number of tokens based on a set of rules. When language is defined as such, it enables the accumulation of knowledge. To form such a system, the segmentation of an external stimulus and internal state, their association, and maintaining signal honesty are considered necessary components.

I hypothesize that proto-speech emerged from the process of mutual segmentation of song string and behavioral contexts (Merker and Okanoya 2007; Okanoya and Merker 2007). Once speech had gained the combinatory property by which new

expressions became possible, the speech signal could now point to nonexistent or imaginary entities. This marked the beginning of imagination. By freely combining concepts that were not associated, humans came to develop their imagination and creative thinking. However, at the same time, this also marked the beginning of manipulative communication, because with language, anything could be expressed without grounding it in the traits of the speaker. This also made language a dishonest signal in the sense of signal honesty (Bradbury and Vehrencamp 2012). Nevertheless, humans continued to use language once it had been acquired evolutionarily. Why is this possible? This consideration would also give shape to extraterrestrial “language.”

11.5.1 *Language-Like Signals: Honest and Dishonest Components*

One of the reasons why language, a dishonest signal, survived could be because language as expressed speech has multiple components. Speech comprises vocal behavior used in face-to-face contexts. This means that speech, in its original mode, is used in real time, in proximity, and together with visual information. Speech behavior includes emotional information such as prosody, facial expression, and bodily movement. This emotional information mostly consists of honest signals, because they cannot be manipulated intentionally (Zuckerman et al. 1979). At the same time, of course, speech content comprises editable information. In face-to-face communication, if the speech content intentionally contained false information, prosodic or facial emotion would convey that the content was untrue. Honesty of speech content was thus guaranteed by honesty of speech behavior (Fig. 11.4). In

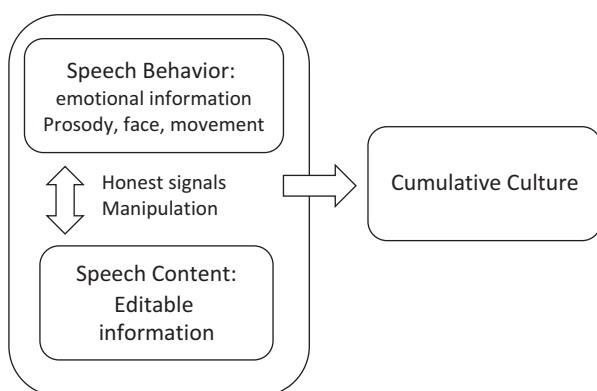


Fig. 11.4 Since the emergence of language, communication content has been divided into linguistic content and emotional information. Since the invention of telecommunication, text content alone is often conveyed, but the accompanying emotional information is often discarded. This situation causes problems in maintaining linguistic communication

this way, human speech was utilized and evolved as a useful tool to accumulate knowledge.

If an extraterrestrial language exists, it should also contain multiple components, some of which should support the accumulation of information and the remainder of which should contribute to securing signal honesty.

11.5.2 The Future of Human Communication

The above scenario might account for the evolution of speech up to the invention of telecommunication in humans. Telecommunication first began with the invention of speech recording by means of non-acoustical, mostly visual notations. Sophisticated visual notations of speech ultimately led to the invention of letters. Because visual notations and letters continued to include emotional information and cost of production, the honesty of the contents was still not entirely violated (Lachmann et al. 2001) and the primary mode of communication continued to be face-to-face. As electrical devices for telecommunication advanced, however, the face-to-face mode of speech communication began to lose its position as the primary mode of communication. In modern society, a great deal of work is conducted through telecommunication devices in which most of the information is text based. We examined how emotional content could be transmitted in telecommunication devices and found that the sense of emotion transmission is very low in text-based communication (Arimoto and Okanoya 2015).

Although text-based communication is efficient in terms of the time, cost, and accuracy of both parties, it lacks the signal honesty necessary for fruitful communication. Additionally, since devices develop much more quickly than a single generation of humans, different generations are imprinted with different means of information transfer (Kelly 2016). Most current social problems are rooted in these simple facts. Now is the time to consider how we should design future means of communication.

11.5.3 Can Solaris Exist?

To revert to the introduction, I described two novels whose theme is language in nonhuman extraterrestrial intelligence. I think Solaris could not exist because it is a single organism that does not require communication and competition with other similar organisms. The planet Solaris does not require segmentation, association, or signal honesty. Thus, it does not require a system of communication, and no self-awareness would evolve in such a planet. Likewise, the heptapod in *Story of Your Life* could not obtain the time-transcending linguistic system. This is because language is based on the token-state association, and this association depends on the temporal co-occurrence of events (Rescorla 1972). Signal honesty is not supported

in heptapod communication because honesty is not judged in time-transcending situations. Of course, I am by no means criticizing these novels; in fact, I love them for the very reason that they trigger my imagination on important questions of what it is to be human.

11.6 Conclusion

In this chapter, I reviewed the literature on the evolution of acoustic communication in animals. I developed a set of hypotheses to account for the emergence of human speech and language in line with the evolution of animal communication. I found that a discontinuity occurred when humans began to use devices for telecommunication, since these remove the emotional information that supports the honesty of linguistic content. I considered that this might change the way humans use language. When this is extended to the language-like communication system of a hypothesized extraterrestrial one, I can suggest at least that the system should contain multiple components to support the contradictory needs of information accumulation and signal honesty. I can also suggest that such a system would need to function on the axis of time to enable the effective association of tokens and states.

Acknowledgments This work was supported by MEXT/JSPS Grant-in-Aid for Scientific Research on Innovative Areas #4903 (Evolinguistics), Grant Number JP17H06380. I would like to thank Dr. M. Takahasi for his insightful comments on an earlier draft.

References

- Arimoto Y, Okanoya K (2015) Mutual emotional understanding in a face-to-face communication environment: how speakers understand and react to listeners' emotion in a game task dialog. *Acoust Sci Technol* 36(4):370–373
- Bergman TJ (2013) Speech-like vocalized lip-smacking in geladas. *Curr Biol* 23(7):R268–R269
- Bradbury JW, Vehrencamp SL (2012) Principles of animal communication. Sinauer Associates, Sunderland
- Briefer E (2012) Vocal expression of emotions in mammals: mechanisms of production and evidence. *J Zool* 288(1):1–20
- Brudzynski SM (2014) Social origin of vocal communication in rodents. In: Witzany G (ed) *Biocommunication of animals*. Springer, Dordrecht, pp 63–79
- Castellucci V, Pinsker H, Kupfermann I, Kandel ER (1970) Neuronal mechanisms of habituation and dishabituation of the gill-withdrawal reflex in *Aplysia*. *Science* 167(3926):1745–1748
- Chiang T (1998) Story of your life. *Stories of your life and others*, 117–178
- Chomsky N (2000) *New horizons in the study of language and mind*. Cambridge University Press, Cambridge MA
- Clarke E, Reichard UH, Zuberbühler K (2006) The syntax and meaning of wild gibbon songs. *PLoS One* 1(1):e73
- Deacon TW (1998) *The symbolic species: the co-evolution of language and the brain*. WW Norton & Company, New York

- Ehret G (2005) Infant rodent ultrasounds—a gate to the understanding of sound communication. *Behav Genet* 35(1):19–29
- Elowson AM, Snowdon CT, Lazaro-Perea C (1998) Infant ‘babbling’ in a nonhuman primate: complex vocal sequences with repeated call types. *Behaviour* 135(5):643–664
- Fitch WT, Hauser MD (2003) Unpacking “honesty”: vertebrate vocal production and the evolution of acoustic signals. In: Simmons AM, Popper AN, Fay RR (eds) *Acoustic communication*. Springer, New York, pp 65–137
- Geissmann T (2002) Duet-splitting and the evolution of gibbon songs. *Biol Rev* 77(1):57–76. <https://doi.org/10.1017/s1464793101005826>
- Ghazanfar AA, Takahashi DY, Mathur N, Fitch WT (2012) Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Curr Biol* 22(13):1176–1182
- Hammerschmidt K, Whelan G, Eichele G, Fischer J (2015) Mice lacking the cerebral cortex develop normal song: insights into the foundations of vocal learning. *Sci Rep* 5:8808
- Inoue Y, Sinun W, Yoshida S, Okanoya K (2012) Male gibbons change the note order according to the behavioral situations. In: *The Evolution of Language: Proceedings of the 9th International Conference (EVOLANG9)*, World Scientific, Kyoto, Japan, p. 458, 13–16 March 2012
- Jarvis E (2006) Selection for and against vocal learning in birds and mammals. *Ornithol Sci* 5(1):5–14
- Kelly K (2016) *The inevitable: understanding the 12 technological forces that will shape our future*
- Lachmann M, Szamado S, Bergstrom CT (2001) Cost and conflict in animal signals and human language. *Proc Natl Acad Sci U S A* 98(23):13189–13194
- Lem S (1961) *Solaris*, trans. Joanna Kilmartin and Steve Cox. *Solaris, the chain of chance, a perfect vacuum* (London: Penguin, 1981):32
- Liu W-C, Wada K, Nottebohm F (2009) Variable food begging calls are harbingers of vocal learning. *PLoS One* 4(6):e5929
- Meinhardt H, Gierer A (2000) Pattern formation by local self-activation and lateral inhibition. *BioEssays* 22(8):753–760
- Merker B, Cox C (1999) Development of the female great call in *Hylobates gabriellae*: a case study. *Folia Primatol* 70(2):97–106
- Merker B, Okanoya K (2007) The natural history of human language: Bridging the gaps without magic. *Emergence of Communication and Language*. Springer-Verlag, London, pp 403–420
- Morton ES (1977) On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am Nat* 111(981):855–869
- Newman JD (2007) Neural circuits underlying crying and cry responding in mammals. *Behav Brain Res* 182(2):155–165
- Oberweger K, Goller F (2001) The metabolic cost of birdsong production. *J Exp Biol* 204(19):3379–3388
- Okanoya K (2017) Sexual communication and domestication may give rise to the signal complexity necessary for the emergence of language: an indication from songbird studies. *Psychon Bull Rev* 24(1):106–110
- Okanoya K, Merker B (2007) Neural substrates for string-context mutual segmentation: a path to human language. *Emergence of Communication and Language*. Springer-Verlag, London, pp 421–434
- Rescorla RA (1972) *A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement, Classical conditioning II: Current research and theory*. Appleton-Century-Crofts, New York
- Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. *Cognition* 70(1):27–52
- Searcy WA, Nowicki S (2005) *The evolution of animal communication: reliability and deception in signaling systems*. Princeton University Press, Princeton
- Skinner BF (1990) *The behavior of organisms: an experimental analysis*. BF Skinner Foundation, New York

- Slater PJB (1983) The study of communication. In: Halliday TRS, Slater PJB (eds) *Communication*, vol 2. Blackwell Scientific Publications, New York, pp 9–42
- Wright J, Leonard ML (2007) *The evolution of begging: competition, cooperation and communication*. Springer Science & Business Media, Berlin
- Zuckerman M, DeFrank RS, Hall JA, Larrance DT, Rosenthal R (1979) Facial and vocal cues of deception and honesty. *J Exp Soc Psychol* 15(4):378–339

Chapter 12

Evolution of Intelligence on the Earth



Mariko Hiraiwa-Hasegawa

Abstract Life started on the earth about 3.8 billion years ago, but the emergence of animals with simple nervous systems had to wait until about 580 million years ago. The brain is an organ to receive various kinds of information, assess them, and make decisions to produce adaptive behavior. The cost of having a large brain is quite high, and evolution of large brains seems to be a rather rare event. There is only one species with a brain the size of which amounts to 2% of its body size, namely, modern humans. A species that is intelligent enough to discover and utilize electromagnetic waves must have large brains, organs to manipulate objects, a means to communicate ideas about the external world, and a method to verify or refute hypotheses about nature. During the evolution of life on earth, at least one such species, modern humans, has evolved. As we do not know of any life forms other than those on our earth, we have only one evolutionary system within which to investigate the possibility of the evolution of intelligent species. Conclusions remain tentative until we have other examples of the evolution of life forms on planets other than our earth.

Keywords Evolution · Brain · Culture · Civilization · Science

12.1 Introduction

Humans continue to seek other intelligent organisms in the universe in order to make contact with them. So far there is no evidence of such existence. It is interesting to discover, once some life form has evolved on a planet, how the organisms evolved to be intelligent enough to discover and utilize electromagnetic waves. We, humans, are the only species on earth that is capable of carrying out science. Evolutionary biology investigates the mechanisms of evolution and the phylogeny of organisms as the results of evolution on earth. However, as we do not know about the evolution of life other than our own, our evolutionary biology is limited to only one sample of an

M. Hiraiwa-Hasegawa (✉)

The Graduate University for Advanced Studies, Hayama-machi, Kanagawa Prefecture, Japan
e-mail: hasegawamk@soken.ac.jp

evolutionary system, namely, ours. In this chapter, despite this severe limitation of examples, I will speculate on the probability of the evolution of intelligent species by looking through the evolution of life on earth, the evolution of modern humans, and the emergence of science-based civilizations in human history.

12.2 The Brain as an Organ for Complex Decision-Making

As of 2004, we have about 1.5 million scientifically named species. However, nobody knows how many species actually exist on the earth. We know that there are many places where scientific exploration is still far from complete, such as the deep ocean or canopies of tropical rain forests, and indeed, new species continue to be discovered and recorded every year. We have not yet grasped the entire breadth of the evolution of biodiversity on the earth. The most recent estimates covering all living organisms, including bacteria, estimate the existence of about 1–6 billion species (Larsen et al. 2017).

The origin of life on the earth is estimated to be about 3.8 billion years ago. Since then, numerous species have evolved and become extinct. It has been said that among all the species that have appeared on the earth so far, 99% of them have already become extinct. Even the extinction of the entire genera or families has occurred rather frequently. Life has been always on the verge of extinction.

In order for any “intellectual” species to evolve from among those organisms that evolved on the earth, the most important event should have been the evolution of the nervous system, which eventually resulted in the evolution of large brains. The nervous system has evolved for an individual animal to deal with its environment: to receive information from its environment, assess it, compare it with other information, and make an optimal decision for the next action.

It all started with the evolution of animals that actively move by themselves in response to the environmental change around them. Plants do not move, so that they do not need such a system to deal with the changes in environment moment by moment: nevertheless they, too, have to respond to the changes in their environment, and they do it in entirely different ways from those of animals. The evolution of animals had to wait until about 580 million years ago in the 3.8 billion years of the history of life on earth.

Nervous systems can be roughly divided into two groups of neurons: a group which receives sensory information and another which sends information for motor output. As animals became more complex, more neurons were added to play an intermediary role between these two groups. Neurons to store memories also have been added and increased, and the more the animal has had to respond with complicated behavior, the larger its brain has become (Kaas 2016).

An intelligent organism must have a relatively large brain compared to its body size. Among the numerous species that exist on the earth, modern humans (*Homo sapiens*) have the largest brain in relation to their body size. The human brain weighs about 1500 grams and accounts for 2% of body weight. There is no other animal

with such a large brain. Chimpanzees and cetaceans have relatively large brains, but their brains account for at most 1% of their body weight. These observations strongly suggest that the emergence of large brains is rather a rare incident in the entire evolution of life.

The reason for this rarity comes from the high cost of developing and maintaining a large brain. Although the human brain accounts for only 2% of body weight, it requires more than 20% of the entire energy intake just to maintain it. Growing such a large organ also requires a lot of energy and a long period of time. For most of the animals living a simple life, the costs of having such a large brain seem to far exceed the benefits accruing therefrom. Animals that can afford to have large brains will necessarily be large-bodied and long-lived species that have plenty of opportunities to exploit the outcome of their large brains.

During the evolutionary history of life on earth, there has been no general tendency for all organisms to develop large brains through time. Plants have no brain, and most invertebrates have only simple nervous systems, but nevertheless they flourish all over the world. There are many different strategies for organisms to survive and reproduce in their environment, and large brains have evolved only in some special cases where the benefits exceed the costs of having them.

Mammals in general have large brains relative to body size, but the primates, the order to which humans belong, have especially large brains. Why do they have large brains? Though there has been a great deal of research and many controversies, there is a consensus among researchers that the most important reason may have been the complexity of social life (Dunbar 1998).

12.3 Social Brain Hypothesis

All the large-brained primates are diurnal and social. They live in a semi-closed social group, identify themselves with each other, and have a social ranking system. They recognize the social relationships among themselves, differentiating affinity among themselves, such as between mother and offspring, among other related individuals, friends, and antagonists. There is competition among individuals over social status, and related individuals and friends sometimes ally to confront a higher-status individual.

In this situation, an individual who could read others' minds and manipulate them would have an advantage over an individual who could not. Tactical deception is defined as manipulation of other individuals' behavior by giving false information or exhibiting ambiguous behavior that can be interpreted multiple ways depending on the context. Byrne and Whiten (1989) explored the frequency of observations of these behaviors among primate species and found that it correlated with the relative brain size but not with ecological parameters like diet composition or territory size. They call it "Machiavellian intelligence," insisting that the driver of the evolution of intelligence among primates was the complexity of social life, not the complexity of ecological settings.

This hypothesis was later expanded to “the social brain” hypothesis. Dunbar (1998) showed that the frequency of tactical deception correlated with the relative volume of the neocortex in the brain, and the relative volume of the neocortex itself correlated with the average size of the stable social group of the species, but not with other ecological parameters. The neocortex is a part of the brain involved in the integration of sensory and motor information, and responsible for decision-making.

Primate social life is a complex of competition and cooperation, and this complexity increases exponentially with the number of individuals constituting the society. Once the brain size of the members of the society starts to increase, it will make up a feed-forward loop, because ever larger brains with better computational ability and memory will yield advantages in dealing with other large brains. The social brain hypothesis explains the evolution of large brains among primates as the results of an evolutionary arms race in social complexity.

Humans evolved as one of these primates. The ancestors of humans had already evolved large brains as an organ to deal with social complexity. In order for the humans to evolve ever larger brains, there must have been an energy source to make that evolution possible. Interestingly, humans have relatively small intestines compared to other primates, calculated from body size (Aiyello and Wheeler 1995). The intestine, as an organ, also requires a lot of energy to maintain; thus, this fact provides considerable insight. Since human ancestors advanced to the savannah from forest, they started to exploit more meat in addition to vegetable food like roots and tubers. At the same time, they are thought to have begun using fires for cooking, which made their food considerably easier to digest, and there opened a way for reduced intestine size. This might have contributed greatly to the use of surplus energy to make their brains larger (Wrangham 2010).

12.4 Environment for Human Evolution

The ancestral line of humans diverged from the lines that led to chimpanzees about 6–7 million years ago. The group *Hominini* means the kinds of apes which habitually adopted upright posture and bipedalism, organisms which appeared at that time in Africa. The oldest fossil hominin, *Sahelanthropus tchadensis*, goes back to 6 million years ago, and there are many others such as *Orrorin tugenensis*, *Ardipithecus ramidus*, and *Australopithecus* spp. thereafter. They ventured into the savannah but continued to utilize forests, preserving the ability to grip branches with digits of hind limbs. The relative brain sizes of these creatures were not especially large but remained about 400 cc, the same level as the current chimpanzee and the gorilla, until about 2.5 million years ago.

Then a new type of Hominin appeared. They had relatively long legs, relatively short arms, and body proportions almost the same as ours. Their bodies were as large as modern humans, and their feet completely adapted to bipedalism, no longer able to grasp branches with hind limb digits. They are classified as genus *Homo*.

Genus *Homo* probably abandoned the life in trees completely and adapted to the life on the savannah, being able to walk and run long distances. This was the time of

the beginning of the ice age, and the environment of Africa became colder and drier, resulting in the reduction of tropical rain forests. In this time of change, though the ancestors of current great apes persisted in the tropical forests, human lines ventured into open grassland.

At the same time, the brain size of genus *Homo* increased to about 800–1000 cc. One of the most famous fossils of early *Homo* is called the Nariokotome boy of *Homo ergaster*, which appeared in east Africa about 1.8 million years ago. It has a body size and structure the same as ours and a brain of about 1000 cc. What was the reason of this sudden increase in brain size? Body size itself had almost doubled, but one of the main reasons might have been related to the adoption of a continuous upright posture and walking. Habitual bipedalism set up a completely different environment: direction of the travel of the body became perpendicular to the body axis, and more information processing would have been needed for the coordination of the upright body.

At the same time, habitual bipedalism completely liberated arms and hands from locomotion. All monkeys and apes have prehensile fingers to grasp and manipulate objects, but, nonetheless, they have to use hands for locomotion as well. Human hands, however, are completely free from the role of locomotion, opening up infinite opportunities to carry things and manipulate things by hand. The opportunities for invention and innovation of tools must have increased tremendously. In addition to that, the increasing opportunities to see the results of manipulative hands of their own would have influenced the understanding of causality of events and of the self as one of the drivers of the causality.

The subsistence of the genus *Homo* also changed dramatically. The primate order was originally adapted to life in forests, eating fruits and leaves. All the great apes continue to lead such lives. However, the genus *Homo* had to survive on the savannah where there was much less rainfall and much less easily obtainable food. On the savannah, proteins are packaged into the shape of large ungulates, but they are hard to catch for onetime frugivores who do not have specialized fangs or claws. The starches are densely accumulated in the shape of roots and tubers underground, but cannot be easily dug up without specialized fingers or snouts. On top of this, the genus *Homo* had to compete for these foods with the expert carnivores and rodents.

This meant that the ancestors of humans who ventured into open grassland had to change their diet, subsistence technology, and social structure. They might well have gone extinct, but somehow they survived. Our ancestors did change their diet and subsistence, not by evolving new bodily parts like fangs and claws but by evolving larger brains. The larger brains enabled them to read each other's mind, share purposes, and cooperate with each other. The social brain already existed as a base.

However, there is a very high hurdle for wide-range cooperation to evolve: individuals in a group must be able to detect and expel the noncooperator among them. If everybody cooperates, everybody can reap good results. However, if there is a noncooperator who enjoys the result without any labor, this strategy will spread in the population, and eventually the cooperative system will collapse. In order for our ancestors to be able to get benefits from cooperation, they had to discriminate cooperators from defectors, and the evolution of social brains was ever more accelerated.

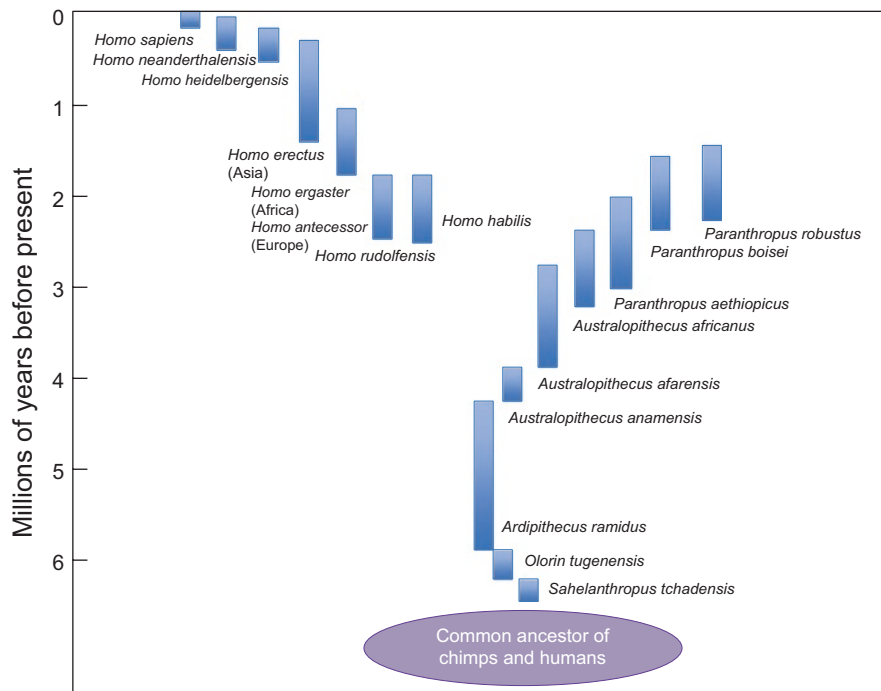


Fig. 12.1 Time line of various Hominins during the Plio-Pleistocene. Duration of existence of each fossil is approximation. The exact phylogenetic relationships among them are unknown

The current fossil evidence suggests that there have been many different species of the genus *Homo* which appeared and disappeared since they first appeared about 2.5 million years ago (Fig. 12.1). *Homo erectus* successfully spread from Africa into the Eurasian continent but eventually went extinct. Neanderthals, who were a distant cousin of our species, evolved around 500 thousand years ago and spread to Europe and the Near East. They had relatively large brains comparable to ours but went extinct by about 30 thousand years ago. Our own species, *Homo sapiens*, evolved in Africa about 200 to 300 thousand years ago and eventually spread to all over the world from 90 thousand years ago onward. There is evidence of hybridization between *Homo sapiens* and *Homo erectus* spp., and *Homo sapiens* and Neanderthals. Also, archeological evidence suggests that, around 90–70 thousand years ago, there was a very severe reduction of population among *Homo sapiens* in Africa. However, in the end, only we *Homo sapiens* survived to this date (Boyd and Silk 2008; Dunbar 2014).

If we look at the history of emergence and extinction of various *Homo* species with large brains as trial and error in the evolution of large brains, we can conclude that the probability of the emergence and continual existence of a large-brained species must be low.

12.5 Language and the Accumulation of Culture

We, humans, use language. We are able to communicate with each other and engage in cooperative action by using language. Through language, we are able to share our knowledge and pass the knowledge onto the next generation. The people of the next generation learn the knowledge and are able to start with that knowledge without the need to discover or invent it by themselves from scratch. Thus, we can revise and accumulate our knowledge. This is our culture. The main reason of our success on this earth is that we have this kind of cumulative culture.

Language is the means to communicate ideas about the world and is essential for human culture. Details of the characteristics of our language and its evolution will be discussed in Chap. 11.

What is culture? In behavioral ecology, culture is defined as a body of information transferred from individual to individual through means other than genetic transmission. With this definition, we can say that nonhuman primates and other animals have cultures of their own, and we can compare them with human cultures to extract characteristics unique to human culture. Chimpanzees, our closest relatives, exhibit various different traditions of food choice, tool using and greeting gestures, which seem to have arbitrarily arisen in a particular society and been culturally transmitted among them over generations (Whiten et al. 1999). Cultural behavior is, thus, not unique to humans, but it seems that we are the only species that has cumulative cultures which are improved through time by adding discoveries and inventions.

Evidence of culture of fossil humans is restricted to some tools made of stone and other enduring materials (Shea 2016). The oldest stone tools are the ones called Oldowan, from about 2.5 million years ago. These tools are stones with edges made by hitting stones against each other and have no characteristic type. This suggests that the makers of these tools did not share the idea of how to make tools but that each individual made the tool by trial and error in his or her own way. Microscopic studies on the surfaces of these tools have revealed that they were used for cutting meat from bones.

From about 1.8 million years ago, we see another type of stone tools called Acheulean hand axes. They have a typical teardrop shape in common, which suggests that the makers of the tools shared the idea of how to make them, and some kind of teaching might have been involved. Acheuleans are said to have been made by *Homo erectus*.

Amazingly there has been no detectable innovation in the making of Acheulean hand axes for nearly 1 million years. But hand axes were not the only tools *Homo erectus* had. There are archeological remains of very sophisticated wooden spears made by *Homo erectus* from about 500 thousand years ago, and there must have been various other tools made from soft materials that didn't last.

The number of tools differentiated for specific purposes began to increase from about 250 thousand years ago and increased exponentially since then. An explosion of diverse art forms like cave paintings, sculptures of figurines, and musical instruments started from about 50 thousand years ago. However, there was still a long way to go before the emergence of civilizations.

12.6 Emergence of Civilizations Based on Science

Modern humans, *Homo sapiens*, emerged about 200 thousand years ago and survived for a long time thereafter with hunting and gathering as their basic subsistence technology. They invented various kinds of tools and artistic objects. However, until the invention of agriculture and domestication of animals, humans remained just another species of great apes. Their population size remained small, and their ecological influence must have been minimal, because the populations with foraging lifestyles were controlled by the availability of foods.

Subsistence based on agriculture and livestock drastically changed the situation: food procurement became stabilized, storing of surplus foods became possible, and sedentary lifestyles began. These changes led to an increase in population size and subsequently generated civilizations and inequality among people. Inequality among people opened up a way for the emergence of a class of people who were able to spend leisurely time “thinking” (Henry 1989).

The invention of agriculture and domestication of animals started about 10 thousand years ago. After that, there appeared a number of civilizations on the earth. Egypt, Mesopotamia, China, and India were called the world’s four great ancient civilizations, but other kinds of civilizations also appeared in Asian areas other than China, including Japan, and in Mesoamerica, South America, and Africa as well (Fernandez-Almesto 2001).

In all of these civilizations, there were people who contemplated the origins of the universe, life, and human beings. Many of those attempts ended up in mythic or religious explanations, but some of them invented devices such as observatories to measure natural phenomena and engaged in objective, potentially “scientific” observation of nature.

Modern humans are equipped with the ability to think logically and the language to communicate logical ideas. It seems that modern humans feel more comfortable when a proper explanation has been provided for a natural phenomenon than otherwise. Therefore, we can assume that there have always been a group of people who were interested in the pursuit of truth about nature in any civilization, but whether their ideas could have become dominant among their entire society depended on numerous other social, economic, and political conditions.

China and India, for example, also invented methods to explore nature objectively, and some of their inventions played an important role in the history of science. However, there is only one civilization that adopted the scientific way of thinking as its basis and ignited the subsequent development of scientific knowledge and science-based technologies. That is the European civilization from the seventeenth century on.

An attempt to explain the world without resort to supernatural power is thought to have germinated in ancient Greece. Thales of Miletus is one of the most famous philosophers who started the tradition. Many of the Pre-Socratic philosophers followed him in explaining nature by constructing theories and hypotheses based only on natural materials and laws.

However, the way toward the establishment of modern science was not easy and straight thereafter. The Roman empire, after the demise of ancient Greece, did not pay much respect to philosophy, and the center of the “scientific way of thinking”

moved to the Islamic world during the medieval ages. Islamic science flourished from about the eighth to the twelfth centuries and played a critical role in preserving Greek philosophy, especially scientific ideas, during the period when the tradition was lost in Europe. But they eventually abandoned this pursuit and returned to religious fundamentalism.

Current scientific activities are based on the following three ideas: (1) there are universal laws and rules that govern natural phenomena; (2) natural phenomena can be logically explained based on those laws and rules; and (3) the explanations should be evaluated by empirical evidence. These principles of scientific methodology were not established overnight but have been slowly built up through the seventeenth to nineteenth centuries. There must have been numerous social and economic conditions that supported the civilization-wide development of science in Europe. Separation of church and state, adoption of democracy, development of a liberal market economy, to name a few, may have played important roles (Ferguson 2012; Morris 2014). Once modern science was established, people in any country or culture have the potential to understand and contribute to science. That is what happened in this world. However, the probability that scientific thought will become the foundation of a civilization may be low.

An intelligent species able to exploit electromagnetic waves must discover electromagnetic waves in the first place. That species must have large bodies equipped with large brains and must be long-lived. As well, that species must engage in scientific pursuits. In order to do this, the species must have some means to share abstract representations and also the means to determine if those representations are true or false. In addition to this, the species must have a large population including individuals who can afford to spend their time on research.

During the 3.8 billion years of history of life on earth, there is only one *Hominin* line that evolved large brains weighing more than 2% of their body weight. Neanderthals had large brains equivalent to ours, but nevertheless they went extinct. Since the emergence of *Homo sapiens* 200 thousand years ago, various civilizations have flourished in different parts of the earth, but only one civilization established science-based societies. The probability of the emergence of an intelligent species that exploits electromagnetic wave seems to be low.

However, it did occur once on our earth in our history of life of 3.8 billion years. There are numerous habitable planets in this universe, and it is possible that other intelligent organisms evolved somewhere. We are not sure whether we can have contact with them while our civilization still lasts, and it is an entirely different issue whether the contact will be a happy one for each other or not.

12.7 Conclusion

When we consider life on earth as an example, we can say that, in order for an intelligent species to evolve, it is necessary that the species have a high-level information-processing system like a large brain, manipulative organs like human hands, and a trustworthy communication means like human language. In the entire 3.8 billion

years of history of life on earth, we are the only species that has evolved to be able to invent science-based civilizations. Based on this fact, we can say that the evolution of intelligent life form is quite rare and could not be achieved without billions and billions of trials and errors. If there is a planet suitable for the evolution of any life form, and if we assume that reproduction with modification through competition and selection is a universal law for any life form, it will inevitably lead to the evolution of animals with large brains. However, we have only one sample of evolutionary systems, namely, the one on this earth, so the conclusion must be tentative until we have another example to compare with our system.

References

- Aiyello LC, Wheeler P (1995) The expensive hypothesis: the brain and the digestive system in human and primate evolution. *Curr Anthropol* 36:199–221
- Boyd R, Silk JB (2008) *How humans evolved*. W. W. Norton & Co, New York
- Byrne RW, Whiten A (eds) (1989) *Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes, and humans*. Oxford University Press, Oxford
- Dunbar RIM (1998) The social brain hypothesis. *Evol Anthropol* 6:178–190
- Dunbar RIM (2014) *Human evolution*. Penguin, New York
- Ferguson N (2012) *Civilization: the six killer apps of western power*. Penguin, New York
- Fernandez-Almesto F (2001) *Civilizations: culture, ambition, and the transformation of nature*. Free Press, New York
- Henry D (1989) *From foraging to agriculture: the Levant and the end of the ice age*. University of Pennsylvania Press, Philadelphia
- Kaas J (ed) (2016) *Evolution of nervous systems*, vol 1–4, 2nd edn. Academic, New York
- Larsen BB, Miller EC, Rhodes MK, Wiens JJ (2017) Inordinate fondness multiplied and redistributed: the number of species on earth and the new pie of life. *Q Rev Biol* 92:229–265
- Morris I (2014) *The measure of civilization*. Princeton University Press, Princeton, NJ
- Shea JJ (2016) *Stone tools in human evolution: behavioral differences among technological Primates*. Cambridge University Press, New York
- Whiten A, Goodall J, McGrew WC, Nishida T, Reynolds V, Sugiyama Y, Tutin CEG, Wrangham RW, Boesch C (1999) Cultures in chimpanzees. *Nature* 399:682–685
- Wrangham RW (2010) *Catching fire: how cooking made us human*. Profile Books, London

Part IV
History of the Earth Reveiled from
Geology

Chapter 13

Formation of Planetary Systems



Shigeru Ida

Abstract The planet formation process regulates the planetary surface environment during the early phase when life may emerge. The frequency and diversity of “habitable planets” are predicted by the planet formation model. Here we review our current understanding of planet formation for our Solar system and general exoplanetary systems. The discovery of diverse exoplanetary systems now requires significant revisions of the classical standard planet formation model that was built to explain our Solar system. A new ingredient, orbital migration of planets, is drastically changing the model. A new idea, pebble accretion, might change the very basis of the model. We also comment on the formation of the magma ocean and early atmosphere and delivery mechanisms of H₂O, C, and N to planets in habitable zones.

Keywords Solar system · Exoplanets · Habitable planets

13.1 Introduction

The frequency and diversity of “habitable” planets in our galaxy is one of the most important problems in the modern astronomy and planetary sciences. Although the planetary conditions for habitability are not clear yet, recent observations have revealed that Earth-size or slightly larger planets (super-Earths/Neptunes) may commonly exist in “habitable zones” in exoplanetary systems (e.g., Petigura et al. 2013), where “habitable zone” is defined as a range of orbital radius where stellar radiation is appropriate for liquid water to be able to exist on the planetary surface under sufficient atmospheric pressure (e.g., Kasting et al. 1993).

In fact, it was recently announced that Earth-size planets were discovered in habitable zones around the stars, Proxima Centauri and Trappist-1. Proxima Centauri is one of the closest (4.25 light years) stars to our Solar system, and three

S. Ida (✉)
ELSI, Tokyo Institute of Technology, Tokyo, Japan
e-mail: ida@elsi.jp

potentially habitable planets were identified for Trappist-1. However, note that both Proxima Centauri and Trappist-1 are red dwarfs (M-type stars) of ~ 0.1 solar mass and ~ 0.001 solar luminosity. Due to the proximity of the habitable zones to the host stars according to the low luminosity, the surface environments of planets in the habitable zones would be very different from that of the Earth. For example, planetary spin would be locked to the orbital rotation. Radiation from the host stars is infrared rather than visible. The amount of ocean may also be different. While super-Earths are also discovered in habitable zones around K-type stars (0.5–0.8 solar mass), Earth-size planets around solar-type stars are still difficult to detect with the current observation resolution (see Chap. 28).

The sizes and orbits of planets are important factors for their habitability. However, surface environments such as atmospheric compositions and amount, climate, ocean, magmatic activities, and asteroidal impacts are direct factors for habitability and are not uniquely determined by the planetary size and orbit. The thermal states of the interior and surface environments during the early phase of the planets when life might emerge depend on the planet formation processes and the architecture of whole planetary systems.

The discovery of diverse exoplanetary systems that have very different architecture from our own Solar system requires reconstruction of planet formation models. About 1% of solar-type stars have Jupiter-mass planets orbiting in the proximity ($< \sim 0.1$ au, where au is the unit of length equal to the mean orbital radius of the Earth, which is $\sim 1.5 \times 10^{13}$ cm) of the host stars, which are called “hot Jupiters.” More stars have Jupiter-mass planets on eccentric orbits (“eccentric Jupiters”). These planets are very different from Jupiter and Saturn in our Solar system, which have almost circular orbits at 5.2 au and 9.6 au. Some of the hot Jupiters even have orbits retrograde to the spin rotation of host stars.

While there is no planetary body inside Mercury’s orbit at 0.4 au in the Solar system, compact systems of close-in super-Earths/Earths or sub-Neptunes are very common in exoplanetary systems (see Fig. 13.1). More than 50% of solar-type stars may have these planets (Mayor et al. 2011). The frequency of these planets could be higher around lower-mass stars.

These architectures of exoplanetary systems imply that dynamical processes, such as orbital migration due to disk-planet interactions and orbital instability associated with planet-planet strong scattering, are important factors in planet formation, although such processes were not considered previously in the classical formation model of the Solar system. From the data of exoplanetary systems, it has come to light that the formation processes of our Solar system are still imperfectly understood. Here I review our current understanding of the processes of planet formation, focusing mainly on our own Solar system. I also comment on the magma ocean that affects the planetary surface environments and transport mechanisms of H_2O , C, and N to planets in habitable zones during planet formation.

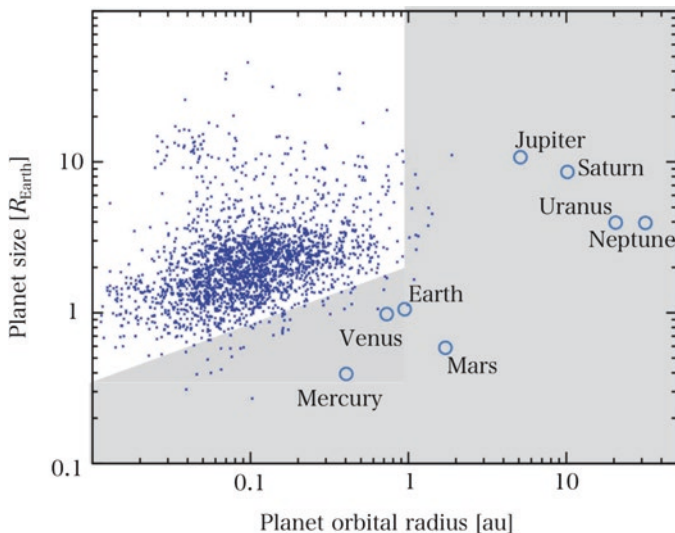


Fig. 13.1 Confirmed planets detected by Kepler space telescope with transit observation. Solar system planets are also plotted. The shaded region represents parameter range beyond detection limit by Kepler observation

13.2 Classical Model and Its Problems

The classical standard model of planet formation was built in the 1970s–1980s (e.g., Safronov 1969; Hayashi et al. 1985) based on disk and planetesimal hypotheses. The model is as follows (also see Fig. 13.2). (1) The remnants of star formation form a protoplanetary disk around the host star. (2) The disk consists of H/He gas by ~99 wt.%, which is the same as the host star. However, the disk temperature is much lower than the star, submicron-sized silicate/iron/icy grains condense in the disk, and 1–10 km sized planetesimals are formed from the grains. (3) Planetesimals grow through collisional coalescence to form terrestrial planets and cores of Jovian planets. (4) In the outer disk regions, because icy grains also condense in addition to silicate grains, cores large enough to cause the onset of runaway gas accretion from the disk are formed, resulting in Jovian planets such as Jupiter and Saturn. (5) In the outermost disk regions, the core growth is so slow that the cores fail to start runaway gas accretion before the disk gas depletion and they are left as mid-sized icy planets such as Uranus and Neptune. Then, the Solar system is completed.

More details of the planetesimal accretion process (step 3) in the classical model, which neglects orbital migration, are as follows. In the early stages, the planetesimal accretion is “runaway growth.” Larger planetesimals grow more rapidly than the others (e.g., Wetherill and Stewart 1989; Kokubo and Ida 1995). In the later stages, the runaway bodies compete with each other, while most planetesimals remain

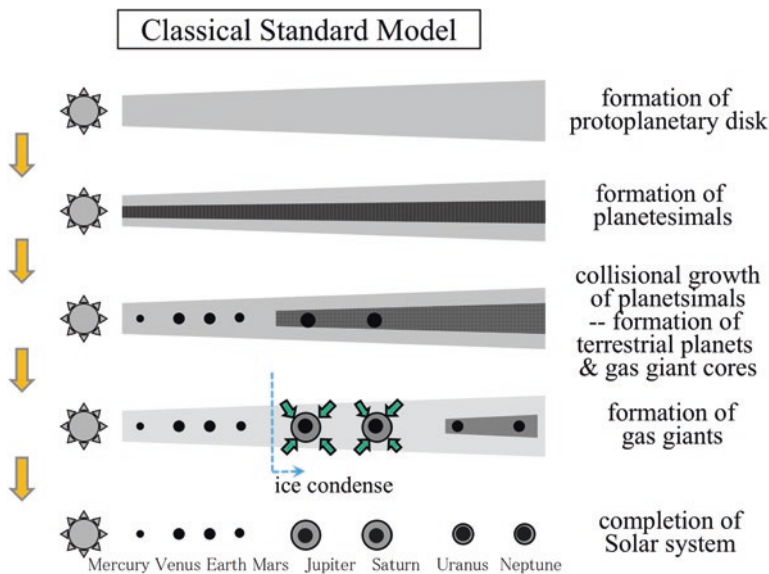


Fig. 13.2 Classical standard model for Solar system formation. For details, see the main text

small (“oligarchic growth,” Kokubo and Ida 1998). The formation timescale of Mars-size bodies is \sim a few $\times 10^5$ years at 1 au for the disk model inferred from the Solar system (Hayashi 1981). The final mass of the oligarchs in the presence of disk gas, called “isolation mass,” is $\sim 0.15 M_E$ at 1 au, where M_E is the Earth mass $\sim 6 \times 10^{27}$ g (Kokubo and Ida 1998) for the pre-solar system disk and it increases with orbital radius. The next stage of growth is via collisions between oligarchs, termed “giant impacts,” which occur after the disk gas depletion (the disk gas strongly circularizes the orbits to inhibit the giant impacts). Mercury and Mars could be oligarchs that avoided collisions during the giant impact stage, which is supported by Hf-W chronology (e.g., Kleine et al. 2009). By a final giant impact to the proto-Earth, the Moon would have been formed (e.g., Canup and Asphaug 2001).

Thus, the classical model seems to beautifully explain the architecture of our Solar system. However, as stated below, it was recently recognized that there are many aspects in the Solar system that cannot be explained by the classical model. It is apparent that the classical model cannot explain the very different architecture discovered in diverse exoplanetary systems.

It has been recognized that the formation of planetesimals from submicron dust grains (step 2) cannot be attained by a simple process. The pairwise growth of dust grains is expected to stall at mm or cm sizes due to bouncing and fragmentation (e.g., Blum and Wurm 2008; Brauer et al. 2008). Furthermore, the “meter-size” barrier that interferes with the further enlargement of the planetesimals (Adachi et al. 1976; Weidenschilling 1977), which has been recognized for a long time, has not been solved yet. Since the disk gas rotates at sub-Keplerian velocities due to pressure

support, the dust particles feel a headwind, and it removes their angular momentum to result in inward migration of the particles. The migration timescale for meter-sized bodies is only ~ 100 years at 1 au. According to this difficulty, a new model, “pebble accretion,” has been proposed (Sect. 13.3).

The formation of a large enough core for Jupiter within a typical disk lifetime \sim a few million years (e.g., Haisch et al. 2001) is also a long-standing issue. The mass distribution of planets in the Solar system is very different between the inner and outer regions that are divided by the asteroid belt. The inner planets are low-mass and rocky, while in the outer regions, planets are much more massive and consist of a large amount of H/He gas. The difference is often attributed to a “snow line” at ~ 3 au. Beyond the snow line, because icy grains condense, more solid materials are available for more massive planetary cores. For the onset of runaway gas accretion, a core mass larger than $\sim 10 M_E$ is required (e.g., Bodenheimer and Pollack 1986; Ikoma et al. 2000). However, the classical model predicts a longer accretion timescale than a typical disk’s lifetime for a $10 M_E$ core at ~ 5 au. Furthermore, the core isolation mass (the maximum protoplanet mass) predicted by the oligarchic growth model is lower than $10 M_E$.

The oligarchic growth model predicts that the final planet mass continuously increases with the orbital radius (at the snow line, the predicted mass jumps by a factor of a few). However, Mars at 1.5 au is 10 times less massive than the Earth at 1 au, and there is no planetary-mass body in the asteroid belt at $\sim 2\text{--}3$ au. Even with the strong perturbations of Jupiter, N-body simulations from a continuous planetesimal distribution cannot reproduce such a small Mars analog (Chambers 2001). The classical model cannot explain the deficit of planetary bodies inside Mercury’s orbit at 0.4 au, either.

Another problem is the two distinct isotopic groups in asteroids: C-type (carbonaceous) and S-type (non-carbonaceous) bodies, which are obtained by the analysis of carbonaceous and non-carbonaceous meteorites. Although their radial distributions somehow overlap, their stable isotope ratios are very different (Kruijer et al. 2017). The isotope ratios are usually interpreted as reflecting the birth locations of solid materials in the disk, so that the different isotope ratios between carbonaceous and non-carbonaceous meteorites are inconsistent with the current radial distribution of C-type and S-type asteroids.

13.3 Pebble Accretion

In Sect. 13.2, the “meter-size barrier” for grain growth was mentioned. Due to aerodynamic gas drag, small particle motions are coupled to the gas. The condensed particles grow through pairwise collisions in the disk. While the growth timescale (the mass-doubling timescale) is almost independent of the particle mass, the coupling between the particle and the gas becomes weaker, and the migration timescale becomes shorter as the particle grows. When the particles grow to ~ 10 cm size, the migration dominates the growth, and the migration of the particles actually starts.

The particles never reach planetesimal sizes before falling onto the star. This is called the “meter-size barrier” or “drift barrier” (Weidenschilling 1977). Although it is pointed out that fluffy ice grains may grow sufficiently rapidly (Okuzumi et al. 2012; Kataoka et al. 2013), the drift barrier is considered to be still a serious problem.

In the past, it was assumed that particles settle into a thin midplane until they collapse into clumps from their self-gravitational force (Goldreich and Ward 1973). However, the particle layer will be stirred up by the instability induced by the sedimentation that is called Kelvin-Helmholtz instability, which prevents the collapse (Weidenschilling 1995; Sekiya 1998). One possible process that leads to the collapse is “streaming instability” (Youdin and Goodman 2005). In this process, when pebbles form aggregates, the aggregates move more slowly in the gas by their inertia, and they accumulate more surrounding pebbles that move faster to eventually form gravitationally bound clumps. However, in order for the streaming instability to occur, particle sizes must be larger than cm sizes and the pebble-to-gas mass ratio must be larger than ~ 0.02 (Carrera et al. 2015), which is difficult to be realized because pebbles migrate so fast (Krijt et al. 2016; Ida and Guillot 2016). It is not yet clear if the streaming instability can form planetesimals in protoplanetary disks. Clumps may form at local discontinuities of the protoplanetary disk where the radial pressure gradient is positive and migration of pebbles will be halted (e.g., Johansen et al. 2014).

When icy pebbles of ~ 10 cm sizes pass the snow line, many small silicate grains would be ejected during sublimation of icy components (Saito and Sirono 2011; Morbidelli et al. 2016). Because the small silicate grains are coupled to the gas due to the strong gas drag, their migration is very slow. Together with rapid migration of icy pebble, the ejected silicate particles pile up just inside the snow line. This pileup could lead to rocky (silicate) planetesimal formation (Ida and Guillot 2016; Drazkowska and Alibert 2017). The inefficient accretion of small particles might also be important for the dichotomy of the Solar system (Sect. 13.4).

Once 100–1000 km planetesimals (seed planets) are formed, their motions are decoupled from the disk gas. The pebble’s motion deviates from its original orbital trajectory due to the planetesimal gravity, increasing gas drag and enhancing the cross-section of planetesimals (Ormel and Klahr 2010, Lambrechts and Johansen 2012). The capture cross-sections of the planetesimals for pebbles become very large, comparable to the Hill sphere for the large planetesimals. While the formation mechanism of planetesimals with >100 km size is unclear, pebble accretion can be very efficient and potentially solves the problem of core accretion for gas giants within the disk lifetime (Lambrechts and Johansen 2012).

Growth of pebbles from small grains and onset of their migration are earlier in an inner region, and the pebble formation front propagates outward (e.g., Sato et al. 2016). Because pebbles are still formed in the outer regions of the disk and migrate all the way through the disk, the planets can grow without any limit. This is very different from the idea of isolation mass in the oligarchic growth model where a protoplanet accretes nearby planetesimals in a limited range of radial distance from the protoplanet orbit (Figs. 13.3 and 13.4). When a planet becomes large enough to

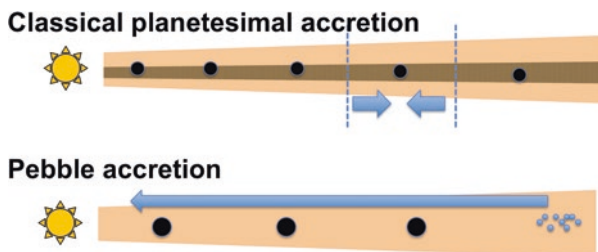


Fig. 13.3 Schematic illustration of the classical planetesimal accretion model and pebble accretion model. While planets accrete from nearby planetesimals in the classical model, pebbles are formed in the outer regions of the disk and migrate all the way through the disk

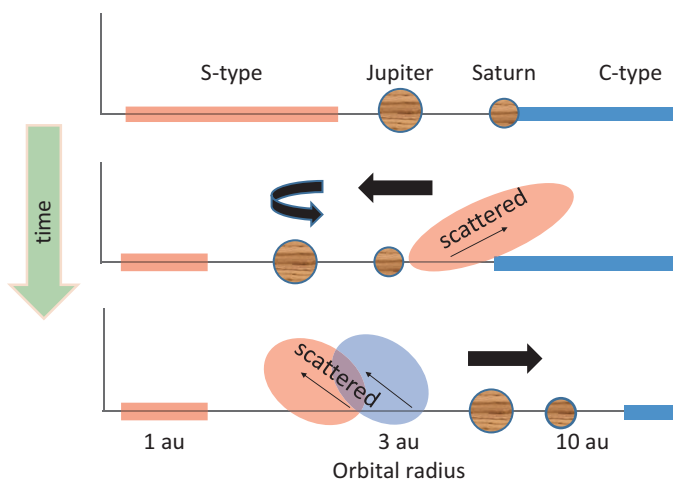


Fig. 13.4 Schematic illustration of “Grand Tack” scenario. Time evolution is from the upper panel to the lower panel. (Based on Kevin Walsh HP <http://www.boulder.swri.edu/~kwalsh/GrandTack.html>)

make a dip in the radial distribution of the gas surface density, the positive radial pressure associated with the dip inhibits migration of pebbles due to aerodynamic gas drag. The dip is created when the planet mass exceeds \sim several M_E at 1 au (Lambrechts et al. 2014). Due to this critical mass being much larger than the masses of the current terrestrial planets in our Solar system, the pebble flux must be truncated by consumption of all the solid materials in the disk (Ida and Guillot 2016) or by formation of a large planet in an outer region (Morbidelli et al. 2016) if pebble accretion is assumed. Pebble accretion is so fast that the truncation at ~ 0.1 – $1 M_E$ to fit the terrestrial planets in the Solar system may need very fine tuning. While giant impact phase is consistent with the classical oligarchic growth model, it would not be consistent with pebble accretion model.

It is often inferred that oxygen isotope ratios reflect the birthplaces of the building block materials. If pebbles were formed in outer regions and migrate all through

the disk, all the bodies formed by pebble accretion would have similar isotope data. However, the observed oxygen isotope ratios among Earth, Mars, meteorites, and comets are clearly different. Further examination is needed to explain the observed variations of oxygen isotope ratios.

As we have discussed so far, pebble accretion model can avoid a serious “meter-size barrier” problem and account for rapid formation of Jupiter. It might also explain the dichotomy between inner and outer regions of the Solar system. However, formation of seed planets has not been clarified. The consistency with giant impacts and diverse oxygen ratios is also a problem.

13.4 Gas Giant Formation

The classical model for formation of gas giants is called “core accretion” (e.g., Mizuno 1980; Stevenson 1982; Bodenheimer and Pollack 1986). The mass of gravitationally bound gas envelope increases as the planet grows. At some critical core mass, the pressure gradient no longer supports the envelope against the planetary gravity, and the envelope starts collapsing along a Kelvin-Helmholtz contraction timescale. As the contraction proceeds, the surrounding disk gas flows onto the planet, rapidly producing a gas giant. The critical mass is typically $\sim 10 M_E$.

Another model is the “disk instability” model (Boss 2000), whereby a self-gravitationally unstable gas disk might fragment into objects that survive as gas giants. While this model could explain wide-orbit gas giants found by direct imaging (e.g., Kalas et al. 2008), it is not easy to explain Jupiter-mass planets at $< \sim 10$ au. We will not discuss the disk instability model any further in this chapter.

With the pebble accretion model, cores with $> 10 M_E$ can easily grow before the disk depletion, which may not be easy in the classical planetesimal model, as mentioned in Sect. 13.2. In the Solar system, the total mass of terrestrial planets that consist mostly of solid materials is only $\sim 2 M_E$, while the total mass of solid materials in the gas planets may be as $> \sim 50 M_E$ in the outer Solar system. It is not clear why such a huge difference in the distribution of solids exists between the inner and outer regions in the Solar system. It is not entirely explained only by the condensation of icy grains beyond the snow line. The pebble accretion model could explain the dichotomy, because accretion of small silicate particles in the regions inside the snow line is less efficient than that of icy pebbles beyond the snow line and because formation of a gas giant which is formed beyond the snow line truncates further pebble flow into the inner regions and stores a large amount of pebbles in the outer regions (Morbidelli et al. 2016).

13.5 Orbital Migration

The classical planet formation model described in Sect. 13.2 assumed in situ growth of planets. However, discovery of close-in Jupiters/Neptunes/super-Earths in exoplanetary systems strongly suggests that planet formation is not necessarily an in situ process. When a planet becomes massive enough, it gravitationally interacts with the protoplanetary disk gas and launches density spiral waves in the disk. Due to the recoil of the density waves, the planet migrates. The torques from inner and outer disk regions are in opposite directions and roughly cancel each other out. In a locally isothermal smooth disk, the outer torque is stronger, and the planet migrates inward (Tanaka et al. 2002). This is called “type I migration.” It was found that an Earth-size planet at 1 au and a core with $10 M_E$ at 5 au migrates inward toward the host star in 10^5 years. This short timescale is a serious problem for the formation of the cores of the giant planets because they must form in the disk of lifetime of a few million years.

Since type I migration is driven by an imbalance between the inner-disk and outer-disk torques, small thermal/dynamical differences of the disk structure can change the speed and even the direction of the migration; in dense (non-isothermal) disk regions, the planet migrates outward (Paardekooper and Papaloizou 2009). However, the range of parameters for outward migration is limited (Baruteau and Masset 2013) and has not been well determined. Furthermore, even if the migration is outward, it is still too fast.

When the planet mass becomes comparable to a Jupiter mass, the planetary gravity is strong enough to open a gap in the disk. The planet is fixed in the gap and migrates inward with the disk gas that spirals in toward the host star. This is called “type II migration” (Lin and Papaloizou 1986). Type II migration is the most popular model for the origin of hot Jupiters (Lin et al. 1996). The disk radially diffuses due to angular momentum transfer by turbulent viscosity. Due to angular momentum per unit mass being proportional to the square root of orbital radius, the conservation of total angular momentum results in inward migration of gas except outermost parts of the disk (e.g., Lynden-Bell and Pringle 1974).

Type I migration speed is proportional to the planet mass. Transition of migration from type I to type II avoids a migration timescale that is too short for massive planets. However, the local diffusion timescale in the disk, which regulates type II migration timescale, is shorter than the global disk lifetime. Type II migration also has the problem of too large speed. Radial velocity observations revealed that most of the Jupiter-mass planets are located at >1 au, which is inconsistent with this theoretical prediction (Hasegawa and Ida 2013). Recent hydrodynamical simulations (e.g., Duffell et al. 2014) suggest that disk gas crosses the gap and a planet in the gap does not migrate together with inward spiral of the disk gas.

Furthermore, in the case of the migration of two giant planets, hydrodynamical simulations (e.g., Masset and Snellgrove 2001) show that type II migration can move outward under certain circumstances. This idea was further developed as the “Grand Tack” model that is discussed in more detail in the next section.

Another migration mechanism of gas giants is planet-planet scattering. While the Solar system is dynamically very stable, the stability condition highly depends on the planetary mass and orbital separation (Chambers et al. 1996; Marzari and Weidenschilling 2002). It is expected that, in some systems, gas giants undergo orbital instability to pump up the orbital eccentricities and even eject other planets, while in the other systems, no instability occurs within the main sequence lifetime of the host stars.

The theoretical modeling shows that the observed distributions of eccentric Jupiters are explained by the planet-planet scattering (Ida et al. 2013). For some fraction of orbital instability, inwardly scattered gas giants approach so close to the host stars that their eccentric orbits are shrunk to close-in circular ones; that is, they are transformed into hot Jupiters (Nagasawa et al. 2008). Since this mechanism can form the discovered retrograde hot Jupiters (the tidal circularization can occur after strong scattering that flips a gas giant's orbit), it is suggested that some of hot Jupiters were formed in this way, rather than by type II migration.

13.6 Formation of Rocky Planets

In Sect. 13.2, we described the scenario of planetesimal accretion neglecting orbital migration. If type I migration is taken into account, planets at >1 au would start migration before they attain the isolation mass (Ida and Lin 2010). While the accretion timescale increases with the planet mass, the type I migration timescale decreases; migration dominates over growth when the planet mass exceeds a threshold value, which is $\sim 0.1 (r/1 \text{ au})^{-1} M_{\text{E}}$ for the proto-solar system disk.

In general, the migration due to the non-isothermal torque is outward in dense gas regions of the inner part of the disk, while it is still inward in the outer disk regions. In this case, planets may converge to the boundary between outward and inward migration regions. It is possible that the converging zone is near 1 au, which may be consistent with the idea that terrestrial planets in our Solar system are formed from the planetesimals originally concentrated at 0.7–1 au (Hansen 2009). This local formation scenario explains why there are no celestial bodies inside Mercury's orbit, why Mars is so small, and why no planet-size bodies exist in the asteroid belt in our Solar system. However, the convergent zone should migrate due to the disk evolution, and it is not clear why close-in super-Earths/Neptunes exist in more than half of the exoplanetary systems. Furthermore, as mentioned before, the range of parameters for outward migration is limited (Baruteau and Masset 2013).

Because gravitational perturbations from the gas giants are strong, they sculpt the orbital architecture of terrestrial planets even if the planets are located far from the gas giants. In the case of the Solar system, the asteroid belt is almost emptied out in the proximity of Jupiter (>3.2 au). The orbital eccentricities and inclinations of

the remaining asteroids are generally very high, but Jupiter's perturbations in its current orbit cannot be responsible for the high eccentricities and inclinations. Something must therefore have happened in the asteroid belt during the Solar system formation stage.

The "Grand Tack" model (Walsh et al. 2011) assumes that Jupiter and Saturn first migrated inward (type II) and then reversed their migration direction, where "tack" stands for turnaround. In a disk-planet system, angular momentum is globally transferred outward, mass is transferred inward, and two adjacent giant planets in a common gap should migrate inward together with disk gas. However, if the outer gas giant (Saturn) is not large enough to make a clear outer edge of the gap, disk gas can enter it from the outer region and continuously cross the gap through interactions with the two giants. The two gas giants can migrate outward, gaining angular momentum from the gap-crossing flow (Masset and Snellgrove 2001). While it is dynamically possible in principle, the disk conditions must be tuned. In our Solar system, if the turnaround ("tack") location is 1.5–2.5 au, the asteroid belt is depleted by Jupiter's perturbations, and it can also explain the small size of Mars. This model can also explain the spectral and isotopic differences between S-type and C-type asteroids.

Another potential way of clearing out the asteroid belt is "sweeping secular resonances" due to progressive disk gas depletion (Ward 1981; Nagasawa et al. 2000). On a relatively long timescale, perturbations from a giant planet on a small body can be approximated by orbit averaging ("secular perturbation"). The orbital eccentricity of the small body oscillates with small amplitude and its arguments of periastron and ascending node circulate. However, if the perturbing planet also suffers from secular perturbations from the disk and the precession periods of both the perturbing body and the small one coincide, the orbital eccentricity of the small body increases to a high value ("secular resonance"). As the gas disk is depleted, the location of the secular resonance migrates, scattering small bodies at different locations one after another. This is called a "sweeping secular resonance," which could also be responsible for the high eccentricities of the asteroid belt (Nagasawa et al. 2000) and clearing the belt by a combination with gas drag (Zheng et al. 2017).

Note that the oscillation amplitude of eccentricity by the secular perturbation is proportional to the eccentricity of the perturbing gas giant. If orbital instability of gas giants occurs and their eccentricities are highly pumped up, the eccentricities of terrestrial planets that are located far inside the gas giant orbit can be pumped up to ~ 1 , and the terrestrial planets are thrown into the host star (Matsumura et al. 2013).

For the Solar system formation, a relatively complicated migration of Jupiter and Saturn, called "Grand Tack" model, has been proposed, as we have pointed out. It potentially solves the problems of small Mars and C-type/S-type asteroids, while it requires a fine-tuning for the initial conditions and disk structure.

13.7 Magma Ocean of Rocky Planets

For planetesimal impacts, the impact velocity is comparable to or slightly larger than an escape velocity from the surface of the planet. The impact velocity is large enough to melt the planetary surface if the planet mass is larger than Mercury (e.g., Safronov 1978; Coradini et al. 1983). An Earth-size planet would melt outer layer into what is called a “magma ocean.”

On the other hand, in the case of pebble accretion, magma oceans may not form. When a planet is embedded in protoplanetary disk gas, it attracts a small amount of the disk gas to form a primordial hydrogen atmosphere (Ikoma and Hori 2012). Since the pebble impact velocity is reduced by the atmospheric gas drag, impact velocity may not be large enough to melt the planetary surface. Pebble accretion heats the atmosphere through the drag rather than the planetary surface by the impact. If giant impacts follow the pebble accretion, a magma ocean may be formed sporadically. Considering that a magma ocean chemically reacts with the atmosphere and there is chemical differentiation in the magma ocean, magma ocean plays a key role in the surface environment of an early planet.

Note that in the classical planetesimal accretion model, their impacts onto the planet not only form a magma ocean but also generate an impact degassing atmosphere, which would be an H₂O or CO₂ gas, oxidizing atmosphere. The present atmospheric masses of the Earth and Venus are $\sim 10^{-4}$ wt.% and $\sim 10^{-2}$ wt.%, respectively. It is expected that the early Earth had a CO₂ atmosphere with a similar amount to Venus.

On the other hand, the primordial hydrogen atmosphere predicted by the pebble accretion model is reducing. The earliest atmosphere is influenced by the dominant accretion mode, the planetesimal accretion, or the pebble accretion. As organic synthesis is generally easier in the reducing atmosphere, this is very important.

In the present terrestrial planets in our Solar system, no hydrogen atmosphere remains. Due to the small weight of hydrogen molecules, the hydrogen atmosphere may have escaped via UV radiation. However, it may have stayed in the atmosphere during the early evolution phase of the planets and on super-Earths in particular, due to their gravity being stronger than Earth-size planets. A simple estimation for the hydrogen atmosphere mass is $\sim 10^{-4} (M_p/M_E)^3$ wt.% where M_p is the planet mass (Stevenson 1982). It is important to clarify how long the hydrogen atmosphere is preserved.

13.8 Tails of Planetesimal Accretion

If planetesimal accretion is considered, the main phase of terrestrial planet accretion would end in 100 Myr. However, impacts of leftover planetesimals should continue although the impact rate would decay with time. From the estimated amount of siderophile elements (iron-loving elements) in the mantle, it has been suggested that

the Earth was impacted by planetesimals with total mass of 1 wt.% of Earth mass after the final core-mantle separation (“late veneer,” the late accretion of asteroidal or cometary material to terrestrial planets), which is often identified as the Moon-forming giant impact (e.g., Tonks and Melosh 1992; Jacobson et al. 2014).

From crater counts on the lunar surfaces, it is suggested that the late impacts onto the Earth-Moon system did not decay monotonically and a late spike existed ~ 3.9 Gyr ago (Tera et al. 1974), which is called the “late heavy bombardment (LHB).” The mechanism for the spike has not yet become clear, although a model of Jupiter-Saturn resonant passing was proposed (Gomes et al. 2005). While the total impact mass during the LHB is much smaller than during the late veneer, the LHB occurred much later than the late veneer may have had a great effect on the birth and evolution of life.

13.9 Water and Organic Molecule Delivery During Planet Formation

The current ocean mass on the Earth is only 0.02 wt.% of the Earth. Even taking into account subsurface water, the total mass may be $< \sim 0.1$ –1 wt.%. The abundance of C and N on the Earth is estimated to be 10^3 and 10^5 times lower than the solar compositions (Lineweaver and Robles 2007). If we consider equilibrium condensation and simple equilibrium temperature that is determined by a balance between stellar radiation heating and blackbody cooling from dust grains, H_2O icy grains condense beyond ~ 3 au. Under 1 atm atmosphere, H_2O condenses on the surface of the Earth at 1 au. However, in the disk gas – that has a density many orders of magnitude lower than 1 atm – H_2O condenses only at a temperature < 150 –170 K. NH_3 and CO_2 grains condense beyond 10 au (CH_4 and CO condense further away). In this sense, it seems reasonable that the Earth is highly depleted in H_2O , C, and N. Then the past discussions were how to deliver H_2O and organic materials from the outer low-temperature regions to the Earth. Impacts of C-type (carbonaceous) asteroids that include H_2O and comets were discussed (e.g., Raymond et al. 2004). Due to the similarity of oxygen isotope ratios between the Earth ocean and carbonaceous meteorites, the impact model of C-type asteroids is supported by many researchers. However, the oxygen isotope ratios of C-type asteroids are diverse, and the similarity in the isotope ratios with the Earth would not be robust.

In the case of pebble accretion, icy pebbles are accreted when the H_2O snow line is located inside of the planetary orbit. Although the current equilibrium temperature is ~ 270 K at 1 au, it is pointed out that the disk temperature at ~ 1 au can be lower than 150–170 K and the H_2O snow line is at < 1 au in optically thick disks of ages ~ 1 Myrs (Oka et al. 2011). In the early phase of the disk, gas accretion through the disk may be so vigorous causing viscous heating that the snow line is in an outer orbit. As the disk accretion decays, the snow line migrates inward, while it goes

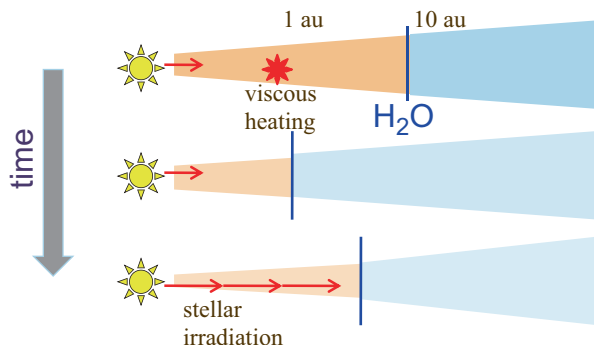


Fig. 13.5 Schematic illustration of migration of the H_2O snow line as the protoplanetary disk evolves

back to ~ 3 au in the final, optically thin phase (Fig. 13.5). During the intermediate phase when the snow line is at < 1 au, icy pebbles formed in the outer regions migrate inward through the disk and are captured by the planet at ~ 1 au. Once the ice is trapped by the planet, they can become gravitationally bound to it even after the snow line goes back to ~ 3 au. Sato et al. (2016), however, showed that the planets at ~ 1 au would capture too much H_2O because the pebble accretion is fast. It may also be a problem that celestial bodies would have similar oxygen isotope ratios in this model, while the observations show that the isotope ratios are different among them in the Solar system.

Even during the intermediate lower-temperature disk evolution phase, the snow lines of NH_3 and CO_2 are still well outside 1 au and the delivery of C and N to the Earth looks difficult. However, it has been suggested that nonequilibrium condensation temperature of HCN is similar to H_2O condensation temperature (e.g., Aikawa et al. 1997). HCN is observationally detected in the protoplanetary disks. Another possible way is that C and N are delivered in refractory forms such as complex organic molecules or carbon grains. Although these refractory carbons are not expected by the equilibrium condensation model, complex organic molecules are found in carbonaceous meteorites. Carbon grains are found in interstellar clouds, interplanetary dust grains, and comets. Since C is a very abundant element in the galaxy, if carbon grains existed in the inner Solar system, the terrestrial planets – including the Earth – should have been full of carbon. However, they are highly depleted in C and N. So, it is a big mystery as to why they have only a tiny fraction of carbon.

If neither H_2O nor C/N molecules are delivered to Earth-size planets located in the so-called habitable zones, the Earth-size planets may not be able to harbor life. It is very important to clarify what controlled the amount of H_2O and C/N in the Earth and how the mechanism can be applied to other exoplanetary systems, in order to discuss habitability of exoplanets.

13.10 Conclusion

The observationally found diversity of exoplanetary systems requires significant revisions of the classical planet formation model. Accordingly, the formation scenario of the Solar system is changing.

High frequency of close-in planets implies orbital migrations due to planet-disk interactions, which play important roles in planet formation processes. However, theoretical models for migrations still remain uncertain. For the Solar system formation, a relatively complicated migration of Jupiter and Saturn, called “Grand Tack” model, has been proposed. It potentially solves the problems of small Mars and C-type/S-type asteroids, while it requires a fine-tuning for the initial conditions and disk structure.

While planetesimal accretion is a basis of the classical model, a new idea, “pebble accretion,” has been proposed. Pebble accretion can avoid a serious “meter-size barrier” problem and account for rapid formation of Jupiter. It might also explain the dichotomy between inner and outer regions of the Solar system. However, it has not been clear how the seed planets accreting small pebbles are formed. Pebble accretion model predicts no giant impact and uniform oxygen ratios among the Solar system bodies, both contradict current understanding. Although we don’t know how much pebble accretion contributes to planet formation relative to the classical planetesimal accretion, if pebble accretion dominates, it will significantly change the view of early Earth surface environments where life would have emerged and evolved.

The dynamical process that delivered H₂O, C, and N to the Earth is still under debate. To discuss a possibility of life on exoplanets in the so-called habitable zones, it is important to clarify the delivery mechanism. We need to elaborate a generalized planet formation model that can consistently explain the Solar system and the diverse exoplanetary systems, to predict surface environments of Earth/super-Earth size exoplanets that are commonly discovered in universe.

References

- Adachi I, Hayashi C, Nakazawa K (1976) The gas drag effect on the elliptical motion of a solid body in the primordial solar nebula. *Prog Theor Phys* 56:1756–1771
- Aikawa Y, Umembayashi T, Nakano T, Miyama S (1997) Evolution of molecular abundance in protoplanetary disks. *Astrophys J* 486:L51–L54
- Baruteau C, Masset F (2013) Recent developments in planet migration theory. *Lecture notes in physics* 861. Springer, Berlin, pp 201–253
- Blum J, Wurm G (2008) The growth mechanisms of macroscopic bodies in protoplanetary disks. *Annu Rev Astron Astrophys* 46:21–56
- Bodenheimer P, Pollack JB (1986) Calculations of the accretion and evolution of giant planets: the effects of solid cores. *Icarus* 67:391–408

- Boss AP (2000) Possible rapid gas giant planet formation in the solar nebula and other protoplanetary disks. *Astrophys J* 536:L101–L104
- Brauer F, Dullemond CP, Henning T (2008) Coagulation, fragmentation and radial motion of solid particles in protoplanetary disks. *Astron Astrophys* 480:859–877
- Canup R, Asphaug E (2001) Origin of the moon in a giant impact near the end of the Earth's formation. *Nature* 412:708–712
- Carrera D, Johansen A, Davies MB (2015) How to form planetesimals from mm-sized chondrules and chondrule aggregates. *Astron Astrophys* 579:A43
- Chambers JE (2001) Making more terrestrial planets. *Icarus* 152:205–224
- Chambers JE, Wetherill GW, Boss AP (1996) The stability of multi-planet systems. *Icarus* 119:261–268
- Coradini A, Federico C, Lanciano P (1983) Earth and Mars-early thermal profiles. *Phys of the Earth and Planetary Int* 31:145
- Drazkowska J, Alibert Y (2017) Planetesimal formation starts at the snow line. *Astron Astrophys* 608:A92
- Duffell PC, Haiman Z, MacFadyen AI, D'Orazio DJ, Farris DB (2014) The migration of gap-opening planets is not locked to viscous disk evolution. *Astrophys J* 792:L10
- Goldreich P, Ward WR (1973) The formation of planetesimals. *Astrophys J* 183:1051–1062
- Gomes R, Levison HF, Tsiganis K, Morbidelli A (2005) Origin of the cataclysmic late heavy bombardment period of the terrestrial planets. *Nature* 435:466–469
- Haisch KE Jr, Lada EA, Lada CJ (2001) Disk frequencies and lifetimes in young clusters. *Astrophys J* 553:L153–L156
- Hansen B (2009) Formation of the terrestrial planets from a narrow annulus. *Astrophys J* 703:1131–1140
- Hasegawa Y, Ida S (2013) Do giant planets survive type II migration? *Astrophys J* 774:146
- Hayashi C (1981) Structure of the solar nebula, growth and decay of magnetic fields and effects of magnetic and turbulent viscosities on the nebula. *Prog Theor Phys Supp* 70:35–53
- Hayashi C, Nakazawa K, Nakagawa Y (1985) Formation of the solar system. In: Black DC, Matthews MS (eds) *Protostars and planets II*. University of Arizona Press, Tucson, pp 100–1153
- Ida S, Guillot T (2016) Formation of dust-rich planetesimals from sublimated pebbles inside of the snow line. *Astron Astrophys* 596:L3
- Ida S, Lin DNC (2010) Toward a deterministic model of planetary formation. IV. Effects of type I migration. *Astrophys J* 719:810–830
- Ida S, Lin DNC, Nagasawa M (2013) Toward a deterministic model of planetary formation. VII. Eccentricity distribution of gas giants. *Astrophys J* 775:42
- Ikoma M, Hori Y (2012) In situ accretion of hydrogen-rich atmospheres on short-period super-Earths: implications for the Kepler-11 planets. *Astrophys J* 753:66
- Ikoma M, Nakazawa K, Emori H (2000) Formation of giant planets: dependences on core accretion rate and grain opacity. *Astrophys J* 537:1013
- Jacobson SA, Morbidelli A, Raymond SN, O'Brien DP, Walsh KJ, Rubie DC (2014) Highly siderophile elements in Earth's mantle as a clock for the moon-forming impact. *Nature* 508:84–87
- Johansen A, Blum J, Tanaka H, Ormel C, Bizzarro M, Rickman H (2014) The multifaceted planetesimal formation process. In: Beuther H, Klessen RS, Dullemond CP, Henning T (eds) *Protostars and planets VI*. University of Arizona Press, Tucson, pp 547–570
- Kalas P, Graham JR, Chiang E et al (2008) Optical images of an exosolar planet 25 light-years from earth. *Science* 322:1345–1348
- Kasting JF, Whitmire DP, Reynolds T (1993) Habitable zones around main sequence stars. *Icarus* 101:108–128
- Kataoka A, Tanaka H, Okuzumi S, Wada K (2013) Fluffy dust forms icy planetesimals by static compression. *Astron Astrophys* 557:L4
- Kleine T et al (2009) Hf–W chronology of the accretion and early evolution of asteroids and terrestrial planets. *Geochim Cosmochim Acta* 73:5150–5188

- Kokubo E, Ida S (1995) Orbital evolution of protoplanets embedded in a swarm of planetesimals. *Icarus* 114:247–257
- Kokubo E, Ida S (1998) Oligarchic growth of protoplanets. *Icarus* 131:171–178
- Krijt S, Ormel CW, Dominik C, Tielens AGGM (2016) A panoptic model for planetesimal formation and pebble delivery. *Astron Astrophys* 586:A20
- Kruijjer TS, Burkhardt C, Budde G et al (2017) Age of Jupiter inferred from the distinct genetics and formation times of meteorites. *PNAS* 1214:6712–6716
- Lambrechts M, Johansen A (2012) Rapid growth of gas-giant cores by pebble accretion. *Astron Astrophys* 544:A32
- Lambrechts M, Johansen A, Morbidelli A (2014) Separating gas-giant and ice-giant planets by halting pebble accretion. *Astron Astrophys* 572:A35
- Lin DNC, Papaloizou JCB (1986) On the tidal interaction between protoplanets and the protoplanetary disk. III – orbital migration of protoplanets. *Astrophys J* 309:846
- Lin DNC, Bodenheimer P, Richardson DC (1996) Orbital migration of the planetary companion of 51 Pegasi to its present location. *Nature* 380:606–607
- Lineweaver CH, Robles J (2007) On the universality of the elemental depletion pattern between a star and its rocky planets. AGU abstract id. P41A-11
- Lynden-Bell D, Pringle JE (1974) The evolution of viscous discs and the origin of the nebular variables. *MNRAS* 168:603–637
- Masset F, Snellgrove M (2001) Reversing type II migration: resonance trapping of a lighter giant protoplanet. *MNRAS* 320:L55–L59
- Marzari F, Weidenschilling SJ (2002) Eccentric extrasolar planets: the jumping Jupiter model. *Icarus* 156:570
- Matsumura S, Ida S, Nagasawa M (2013) Effects of dynamical evolution of giant planets on survival of terrestrial planets. *Astrophys J* 767:129
- Mayor M et al. (2011) The HARPS search for southern extra-solar planets XXXIV. Occurrence, mass distribution and orbital properties of super-earths and Neptune-mass planets. eprint arXiv:1109.2497
- Mizuno H (1980) Formation of the giant planets. *Prog Theor Phys* 64:544–557
- Morbidelli A, Bitsch B, Crida A et al (2016) Fossilized condensation lines in the solar system protoplanetary disk. *Icarus* 267:368–376
- Nagasawa M, Tanaka H, Ida S (2000) Orbital evolution of asteroids during depletion of the solar nebula. *Astron J* 119:1480–1497
- Nagasawa M, Ida S, Bessho T (2008) Formation of hot planets by a combination of planet scattering, tidal circularization, and the Kozai mechanism. *Astrophys J* 678:498
- Oka A, Nakamoto T, Ida S (2011) Evolution of snow line in optically thick protoplanetary disks: effects of water ice opacity and dust grain size. *Astrophys J* 738:141
- Okuzumi S, Tanaka KH, Wada K (2012) Rapid coagulation of porous dust aggregates outside the snow line: a pathway to successful icy planetesimal formation. *Astrophys J* 752:106
- Ormel CW, Klahr H (2010) The effect of gas drag on the growth of protoplanets-analytical expressions for the accretion of small bodies in laminar disks. *Astron Astrophys* 520:A43
- Paardekooper S-J, Papaloizou JCB (2009) On the width and shape of the corotation region for low-mass planets. *MNRAS* 394:2283
- Petigura EA, Howard AW, Marcy GW (2013) Prevalence of Earth-size planets orbiting sun-like stars. *PNAS* 110:19273–19278
- Raymond SN, Quinn T, Lunine JI (2004) Making other earths: dynamical simulations of terrestrial planet formation and water delivery. *Icarus* 168:1–17
- Safronov V (1969) Evolution of the protoplanetary cloud and formation of the earth and planets. Nauka, Moscow
- Safronov VS (1978) The heating of the earth during its formation. *Icarus* 33:35
- Saito E, Sirono S (2011) Planetesimal formation by sublimation. *Astrophys J* 728:20
- Sato T, Okuzumi S, Ida S (2016) On the water delivery to terrestrial embryos by ice pebble accretion. *Astron Astrophys* 589:A15

- Sekiya M (1998) Quasi-equilibrium density distributions of small dust aggregations in the solar nebula. *Icarus* 133:298–309
- Stevenson DJ (1982) Formation of the giant planets. *Planet Space Sci* 30:755–764
- Tanaka H, Takeuchi T, Ward WR (2002) Three-dimensional interaction between a planet and an isothermal gaseous disk. I. Corotation and Lindblad torques and planet migration. *Astrophys J* 565:1257–1274
- Tera F, Papanastassiou DA, Wasserburg GJ (1974) Isotopic evidence for a terminal lunar cataclysm. *Earth Planet Sci Lett* 22:1–21
- Tonks WB, Melosh HJ (1992) Core formation by giant impacts. *Icarus* 100:326
- Walsh KJ, Morbidelli A, Raymond SN, O'Brien DP, Mandell AM (2011) A low mass for Mars from Jupiter's early gas-driven migration. *Nature* 475:206
- Ward WR (1981) Solar nebula dispersal and the stability of the planetary system. I-scanning secular resonance theory. *Icarus* 47:234–264
- Weidenschilling SJ (1977) Accretional evolution of a planetesimals swarm. *MNRAS* 180:57
- Weidenschilling SJ (1995) Can gravitational instability form planetesimals? *Icarus* 116:433–435
- Wetherill GW, Stewart GR (1989) Accumulation of a swarm of small planetesimals. *Icarus* 77:330–357
- Youdin AN, Goodman J (2005) Streaming instabilities in protoplanetary disks. *Astrophys J* 620:459
- Zheng X, Lin DNC, Kouwenhoven MBN (2017) Planetesimal clearing and size-dependent asteroid retention by secular resonance sweeping during the depletion of the solar nebula. *Astrophys J* 836:207

Chapter 14

Evolution of Early Atmosphere



Hidenori Genda

Abstract Earth's early surface environment had great influence on the origin of life through formation of its building blocks. From geological and geochemical evidence, the Earth's atmosphere and oceans appear to have existed since a very early period in the Earth's history. Recent models of planet formation suggest that a significant amount of volatile elements that formed the Earth's atmosphere and oceans was supplied to Earth during its formation. This very early supply of volatile elements is consistent with recent detailed analysis of isotopic compositions of terrestrial and extraterrestrial materials. Chemical equilibrium calculations showed that the volatile elements degassed by accreting bodies contained some reduced gases. Moreover, metallic iron in differentiated bodies accreting on Earth played an important role in reducing the surface environment through producing abundant hydrogen molecules from oceans. These recent studies indicate that the Earth's early atmosphere was more reduced than previously thought, which would be an advantage for formation of the building blocks of life. After Earth's formation, late accretion caused impact erosion and replacement of the pre-existing atmosphere and oceans. However, due to the complicated phenomena of impact erosion, it is difficult to make an accurate estimate of the loss and replacement of the atmosphere at present.

Keywords Volatile supply · Planet formation · Degassing · Redox state · Impact erosion

H. Genda (✉)

Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo, Japan

e-mail: genda@elsi.jp

14.1 Introduction

Our Earth is often called an “aqua planet.” Indeed, 71% of the current Earth’s surface is covered by oceans, which is one of its unique characteristics. However, the mass of the Earth’s oceans is only 0.023 wt% of its total mass, and the mass of the Earth’s atmosphere is two orders of magnitude lower than that of the oceans (e.g., Genda 2016). This small amount of volatile elements on the surface would have played an important role in the emergence, evolution, and prosperity of life on Earth. Regarding the emergence of life, the Earth’s early surface environment had great influence on the formation of life’s building blocks. For example, a redox state of the surface environment (atmosphere and oceans) affects the production efficiency of amino acids (Schlesinger and Miller 1983).

In this chapter, we first address the following question: how early have the Earth’s atmosphere and oceans existed from geological, geochemical, and theoretical points of view? Then we review the supply process and timing of volatile elements on Earth based on recent planet formation theory. We also discuss the redox state of the Earth’s early surface environment, which has been revealed to be more reduced than previously thought. Finally, we discuss the effect of late accretion on the early atmosphere after Earth’s formation.

14.2 Evidence for Atmosphere and Oceans on Early Earth

14.2.1 Atmosphere

The major constituents of current Earth’s atmosphere are N₂ (78%), O₂ (21%), Ar (1%), and CO₂ (0.03%), and its total mass is 5×10^{18} kg. How early has the Earth’s atmosphere existed? Since most elements forming the current Earth’s atmosphere have interacted with oceans and the mantle throughout the Earth’s history, it is difficult to infer the formation age of the atmosphere. However, because noble gases such as Ar are chemically inert, those present in the current Earth’s atmosphere have been present in place ever since they accumulated there.

Hamano and Ozima (1978) estimated the timing and degassing rate of Ar from the Earth’s interior and showed that more than 80% of the internal Ar must have degassed to the surface during the Hadean era (before 4.0 Ga). The present $^{40}\text{Ar}/^{36}\text{Ar}$ ratio in the atmosphere is very low (≈ 295.5), compared to the extremely high $^{40}\text{Ar}/^{36}\text{Ar}$ ratios in mid-ocean ridge basalt samples ($>30,000$), which are derived from the upper mantle (e.g., Ozima and Podosek 2002). The high $^{40}\text{Ar}/^{36}\text{Ar}$ ratio in the mantle is caused by the accumulation of ^{40}Ar via the radioactive decay of ^{40}K in the mantle (half-life 1.25 Gyr). Therefore, a low $^{40}\text{Ar}/^{36}\text{Ar}$ ratio in the atmosphere indicates that Ar would be degassed at the time before ^{40}K decayed. The Ar must have degassed with other volatiles, such as N₂, CO₂, and H₂O.

This early degassing of the Earth's atmosphere is also confirmed more quantitatively from the excess ^{129}Xe in the mantle relative to the atmospheric content, since its excess indicates that the short-lived ^{129}I still existed when degassing of volatile elements from the mantle took place. The half-life of ^{129}I (16.7 Myr) further imposes a quantitative constraint on the time of degassing. The degassing time must be comparable with or shorter than its half-life, suggesting a very early mantle degassing (e.g., Ozima and Podosek 2002).

14.2.2 Oceans

Direct evidence for the existence of an ancient ocean is provided by the existence of ancient pillow lava and sedimentary rocks (see Fig. 14.1) formed throughout the Earth's history and found in various locations. These rocks can only be formed under large bodies of water, providing the evidence of early oceans. The oldest pillow lavas and sedimentary rocks were dated back to 3.7–4.0 Ga (Appel et al. 1998; O'Neil et al. 2008; Maruyama and Komiya 2011; Komiya et al. 2015). Therefore, the Earth's oceans must have existed at 3.7–4.0 Ga.

There are no geological rock records from the Hadean era on Earth (e.g., Bowring and Williams 1999). However, it is possible to locate zircon (ZrSiO_4) that can be dated before 4.0 Ga. Zircon is a mineral that is highly resistant to erosion, weathering, and metamorphism. During the Hadean era, all Earth's rock records have vanished and/or have been destroyed, most likely by early intense meteor bombardment, but zircon grains are thought to have survived this era (e.g., Marchi et al. 2014). A part of Hadean zircon grains has high oxygen isotope ($\delta^{18}\text{O}$) values (Wilde et al. 2001), indicating that the Earth's oceans existed during the Hadean era. High $\delta^{18}\text{O}$ values of zircon grains compared with the mantle are produced by low-temperature

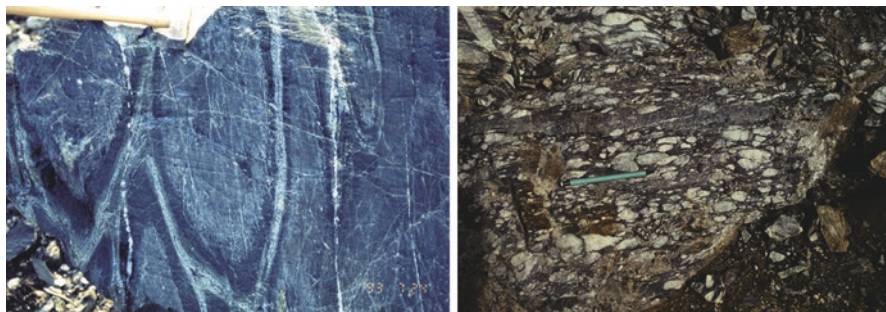


Fig. 14.1 Pillow lava basalt (left) and sedimentary rock (right) in Isua, Greenland. The pillow lava basalt erupted underwater at 3.8 Ga. The gray-blue portion is the core of the pillow and is mantled by a dark blue portion, which in turn is rimmed by pale-colored chilled margins, together with the matrix. The sedimentary rock is a conglomerate interlayered with mafic terrigenous sediments. The scales are given by a hammer at the top of the left photo and a pencil in the middle of the right photo. (Courtesy of Museum of Evolving Earth, Tokyo Institute of Technology)

interactions between rocks and liquid water. Zircon grains with the ages of 3.91–4.28 Ga have high $\delta^{18}\text{O}$ values, suggesting interactions between the continental crust and oceans (Wilde et al. 2001; Mojzsis et al. 2001). Again, this presents evidence for the existence of Earth’s oceans during the Hadean era.

14.3 Origin of Earth’s Atmosphere and Oceans

From geological and geochemical points of view, the Earth’s atmosphere and oceans appear to have existed since a very early moment in the Earth’s history. As shown below, recent models of planet formation suggest that a significant amount of volatile elements that formed the Earth’s atmosphere and oceans was supplied to Earth during its formation.

Planets are formed in a protoplanetary disk—composed of gas and dust—around the sun (Fig. 14.2, also see Chap. 13). Terrestrial planets are made primarily from the dry dust component of the disk. The cores of Jovian planets are made from the

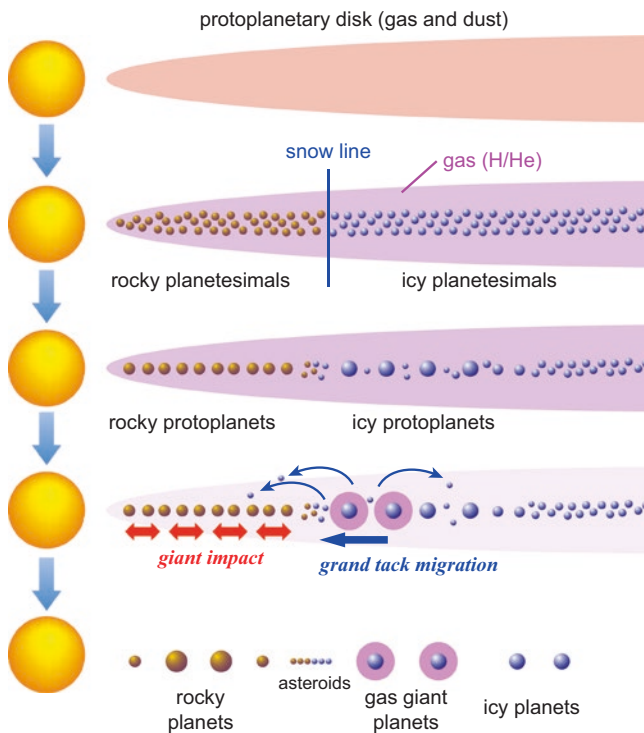


Fig. 14.2 Summary of planet formation in our solar system. A protoplanetary disk composed of gas and dust is formed around the sun during star formation, and planets are finally generated from this disk through several stages

dust component (rock and ice), and subsequently these cores gather the surrounding gas component (H_2 and He) to become gas giant planets like Jupiter and Saturn (e.g., Lissauer and Stevenson 2007). Therefore, the formation of Jupiter and Saturn must be completed before the disk gas dissipates (~ 1 Myr, Natta et al. 2000). Once gas giant planets form, the surrounding small planetesimals are scattered by the strong gravitational influence of these gas giant planets. Some of the planetesimals with volatile elements enter the region of terrestrial planet formation; a few hit rocky protoplanets, which grow into terrestrial planets with a supply of volatile elements on them (Morbidelli et al. 2000; Raymond et al. 2009). Moreover, Jupiter and Saturn possibly migrate inward to the present orbital location of Mars and then return to their current positions, which is called the “Grand Tack” scenario (Walsh et al. 2011, Chap. 13). These migrations stirred the solar system, which lead to the supply of the volatile elements in the Earth’s building blocks (O’Brien et al. 2014).

In the region of terrestrial planet formation, collisions among rocky Mars-sized protoplanets (called giant impacts) take place during or after the dissipation of the disk gas, and this stage lasted for ~ 100 Myrs (e.g., Kokubo and Genda 2010). Therefore, from a theoretical point of view, a significant amount of volatile elements that formed the Earth’s atmosphere and oceans was supplied to Earth during the early stage of its formation. This very early supply of volatile elements to Earth is consistent with recent detailed analysis of isotopic compositions of terrestrial and extraterrestrial materials (Dauphas 2017).

Although the migration of Jupiter and Saturn proposed in the Grand Tack scenario might be thought of as an eccentric idea, it can naturally explain the small mass of Mars relative to Earth and Venus, and the mixing of two isotopically distinct components (carbonaceous and non-carbonaceous meteorites) in the present main asteroid belt (Warren 2011).

Before the 2000s, it was generally thought that all volatile elements were lost by a giant impact that took place during Earth’s formation (Ahrens 1993; Chen and Ahrens 1997). However, through the detailed numerical simulation of the volatile loss by a giant impact, it turned out that a significant amount of volatile elements (especially water) can survive the giant impact stage (Genda and Abe 2003, 2005). This conclusion implies that the volatile elements supplied by Earth’s building blocks may well have persevered and still exist on Earth today.

14.4 Redox State of Early Earth’s Atmosphere

A redox state of the surface environment on early Earth is important for creating the building blocks of life such as amino acids. In the 1950s, Urey and Miller (Urey 1952; Miller 1953) started to conduct experiments to show how the building blocks of life, such as amino acids, could be formed on the early Earth. They used CH_4 , NH_3 , H_2 , and H_2O as the starting materials, which were thought to represent the early Earth’s atmosphere at that time. They found that electric sparks, which imitate

lightning and ultraviolet radiation from the sun on early Earth, stimulated the formation of some amino acids.

However, after their experiments, it was discovered that the Earth's early atmosphere was probably not composed of CH₄, NH₃, and H₂ from the speculation that the volcanic gases are composed mainly of CO₂ and H₂O (Holland 1984). Therefore, the early atmosphere was in a rather oxidized state. Under the oxidized environment (CO₂, N₂, H₂O), experiments similar to the Urey-Miller-type ones were conducted, and the yield of amino acids turned out to be extremely low (Schlesinger and Miller 1983).

According to the planet formation theory discussed before, the Earth's atmosphere and oceans were formed through a supply of volatile elements from small bodies impacting protoplanets and/or very early Earth. An impact degassing of the volatile elements from these colliding small bodies occurred on the surface. Experimental and theoretical studies of impact degassing have shown that incipient and complete dehydration occurs when the impact velocity exceeds about 2 and 4 km/s, respectively (Lange and Ahrens 1982; Tyburczy et al. 1986), whose impact velocity is smaller than the typical impact velocity. The impact velocity is higher than the escape velocity (~5 km/s) for the typical size of protoplanets. Therefore, Earth's early atmosphere and oceans should have been formed through the process of impact degassing, not through the process of volcanic degassing from the mantle.

Recent chemical equilibrium calculations for the impact degassing show that the degassed component has rather reduced chemical species such as H₂, CH₄, and CO, even if the impactor is carbonaceous chondrites, which is the most oxidized group among meteorites (Hashimoto et al. 2007). If the impactor is ordinary or enstatite chondrites, the degassed component should be more reduced (Schaefer and Fegley 2007, 2010). Therefore, these recent studies indicate that the Earth's early atmosphere was more reduced than previously thought.

Moreover, metallic iron ejected from cores of protoplanets during the giant impact stage can re-accrete on the growing Earth (Genda et al. 2017a). Collisions of differentiated bodies after Earth's formation can also spread their metallic iron materials across the early Earth (see Fig. 14.3). If early Earth had oceans, these metallic iron materials extensively reduced water molecules to produce hydrogen molecules in the early atmosphere through the following reaction: $\text{Fe} + \text{H}_2\text{O} \rightarrow \text{FeO} + \text{H}_2$ (Genda et al. 2017b).

The redox state of the surface environment would have changed from the reduced state to an oxidized state through hydrogen escape into space, induced by intensive solar XUV irradiation (e.g., Genda and Ikoma 2008; Kodama et al. 2015) and accumulation of oxidized components to the atmosphere by volcanic activities (e.g., Tajika and Matsui 1992). Since the Hadean era, oxygen fugacity of the Earth's uppermost mantle would have been defined by fayalite–magnetite–quartz buffer, in which oxidized components like H₂O and CO₂ would dominate the magmatic volatile (Trail et al. 2011; Frost et al. 2008). A reduced oxidation state of the Earth's early surface environment would have played an important role in creating the

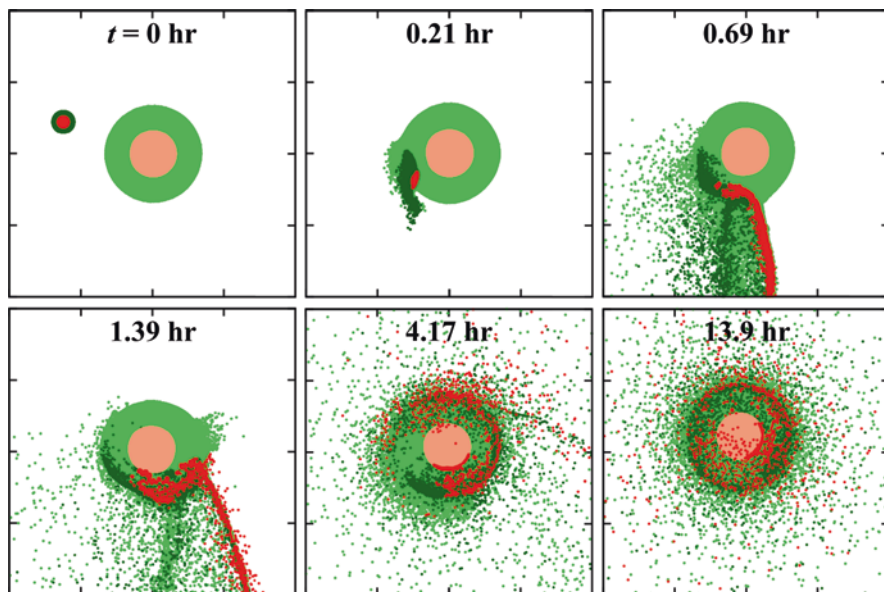


Fig. 14.3 Snapshots for a collision of a large impactor with 1% of the Earth's mass. The impact velocity and impact angle are 16 km/s and 45° . Mantle and core materials for the impactor are colored green and red, respectively, and those for Earth are light green and orange. Some of impactor's core materials have experienced fragmentation and re-accreted on the Earth's surface. The smoothed particle hydrodynamics method (e.g., Fukuzaki et al. 2010) was used for this impact simulation

building blocks of life such as amino acids. In addition, since H_2 and CH_4 helped early Earth to stay warm (Ramirez et al. 2013), the reduced oxidation state of the Earth's early surface environment potentially solves the faint young sun paradox (Sagan and Mullen 1972).

14.5 Atmospheric Erosion by Late Accretion

We can observe many craters on the Moon, which indicates that—after their formation—a lot of bodies hit not only the Moon but also Earth. Figure 14.4 shows the number density of craters >1 km in diameter as a function of the age. Since the cross section of Earth is 20 times larger than that of the Moon, Earth should be exposed to many more impact events. Although the existence of the peak impact flux around 3.5–4.0 Ga—which is called the late heavy bombardment or lunar cataclysm (Tera et al. 1974)—has been debated, the impact flux during the Hadean and early Archean era was much higher than that at present. According to the excess of highly siderophile elements, which often coexist with metallic Fe, in the present Earth's mantle, the total mass of the materials that added to Earth after its formation

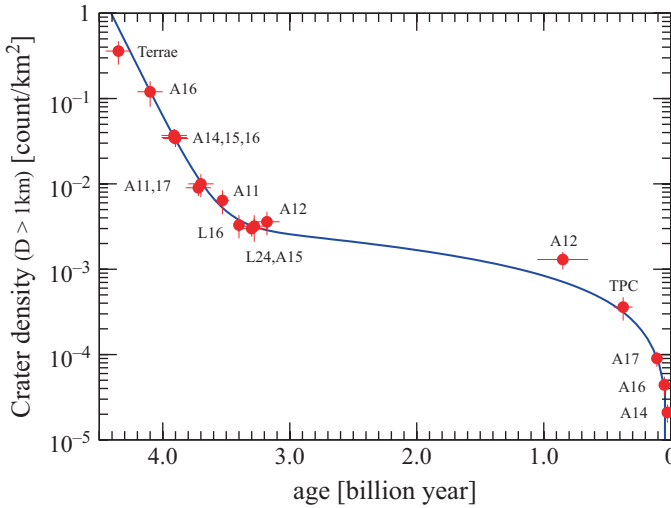


Fig. 14.4 The number density of lunar craters as a function of the age. Only craters with >1 km diameter are considered. Absolute ages were determined from lunar samples returned by Apollo and Luna missions. Moon was formed at ~4.5 Ga. The zero age in the horizontal axis corresponds to the present. A peak in the impact flux may exist around 3.5–4.0 Ga, which is called the late heavy bombardment or lunar cataclysm. (Figure is modified from Neukum and Ivanov 1994)

is estimated to be ~1% of the Earth’s mass (Frank et al. 2016). Therefore, these impacts on Earth would have influenced on the evolution of its early atmosphere.

The collisions of many meteorites with early Earth would have caused atmospheric erosion and replacement of the early atmosphere by volatile elements from impacting bodies. In the 1980s, analytical approach to estimate the effect of the atmospheric erosion was conducted (e.g., Melosh and Vickery 1989), and recently direct numerical simulations have been run. There are two state-of-the-art models of the atmospheric erosion developed by Svetsov (2007) and Shuvalov (2009), which are both based on their results of numerical impact simulations. Many complicated processes are involved in the atmospheric erosion induced by meteorite collision, such as wake formation during the flight of the impactor through the atmosphere, the disruption and deceleration of the impactor, formation of a vapor plume, ejection of impact ejecta, and so on. Due to these complicated processes, completely different results for the evolution of the early atmosphere are obtained by these two erosion models. Extensive loss of the early atmosphere and replacement of the atmospheric compositions are derived from Svetsov’s model, while the atmospheric erosion from Shuvalov’s model is not effective during late accretion. Since collisional outcomes in impact simulations strongly depend on the numerical resolution (Genda et al. 2015, 2017c), much higher resolution simulations are clearly needed to address this issue.

14.6 Conclusions

The Earth's atmosphere and oceans have existed since Hadean era. From a theoretical point of view, a significant amount of volatile elements which formed the Earth's atmosphere and oceans was supplied to Earth during its formation. Impact degassing of volatile elements in some of Earth's building blocks created its early atmosphere (and oceans), whose redox state is rather reduced compared to what was previously thought. Accretion of metallic iron in a differentiated impactor would have played an important role in producing a hydrogen-rich atmosphere on very early Earth. Late accretion after Earth's formation would have also affected the amount and composition of the early atmosphere. However, due to the complicated impact phenomena, it is difficult to make an accurate estimation for the loss and replacement of the atmosphere at present.

References

- Ahrens TJ (1993) Impact erosion of terrestrial planetary atmospheres. *Annu Rev Earth Planet Sci* 21:525–555
- Appel PWU, Fedo CM et al (1998) Recognizable primary volcanic and sedimentary features in a low-strain domain of the highly deformed, oldest known (~3.7–3.8 Gyr) Greenstone Belt, Isua, West Greenland. *Terra Nova* 10:57–62
- Bowring SA, Williams IS (1999) Priscoan (4.00–4.03Ga) orthogneisses from northwestern Canada. *Contrib Mineral Petrol* 134:3–16
- Chen GQ, Ahrens TJ (1997) Erosion of terrestrial planet atmosphere by surface motion after a large impact. *Phys Earth Planet Int* 100:21–26
- Dauphas N (2017) The isotopic nature of the Earth's accreting material through time. *Nature* 541:521–524
- Frank EA, Maier WD et al (2016) Highly siderophile element abundances in Eoarchean komatiite and basalt protoliths. *Contrib Mineral Petrol* 171:29
- Frost DJ, Mann U et al (2008) The redox state of the mantle during and just after core formation. *Philos Trans R Soc A* 366:4315–4337
- Fukuzaki S, Sekine Y et al (2010) Impact-induced N₂ production from ammonium sulfate: implications for the origin and evolution of N₂ in Titan's atmosphere. *Icarus* 209:715–722
- Genda H (2016) Origin of Earth's oceans: an assessment of the total amount, history and supply of water. *Geochem J* 50:27–42
- Genda H, Abe Y (2003) Survival of a proto-atmosphere through the stage of giant impacts: the mechanical aspects. *Icarus* 164:149–162
- Genda H, Abe Y (2005) Enhanced atmospheric loss on protoplanets at the giant impact phase in the presence of oceans. *Nature* 433:842–844
- Genda H, Ikoma M (2008) Origin of the ocean on the earth: early evolution of water D/H in a hydrogen-rich atmosphere. *Icarus* 194:42–52
- Genda H, Fujita T et al (2015) Resolution dependence of disruptive collisions between planetesimals in the gravity regime. *Icarus* 262:58–66
- Genda H, Iizuka T et al (2017a) Ejection of iron-bearing giant-impact fragments and the dynamical and geochemical influence of the fragment re-accretion. *Earth Planet Sci Lett* 470:87–95
- Genda H, Brasser R et al (2017b) The terrestrial late veneer from core disruption of a lunar-sized impactor. *Earth Planet Sci Lett* 480:25–32

- Genda H, Fujita T et al (2017c) Impact erosion model for gravity-dominated planetesimals. *Icarus* 294:234–246
- Hamano Y, Ozima M (1978) Earth-atmosphere evolution model based on Ar isotopic data. In: Alexander EC Jr, Ozima M (eds) *Terrestrial rare gases*. Center for Academic Publ, Tokyo, pp 155–172
- Hashimoto GL, Abe Y et al (2007) The chemical composition of the early terrestrial atmosphere: formation of a reducing atmosphere from CI-like material. *J Geophys Res* 112:E05010
- Holland HD (1984) *The chemical evolution of the atmosphere and oceans*. Princeton Univ. Press, Princeton
- Kodama T, Genda H et al (2015) Rapid water loss can extend the lifetime of planetary habitability. *Astrophys J* 812:165
- Kokubo E, Genda H (2010) Formation of terrestrial planets from protoplanets under a realistic accretion condition. *Astrophys J Lett* 714:L21–L25
- Komiya T, Yamamoto S et al (2015) Geology of the Eoarchean, >3.95 Ga, Nulliak supracrustal rocks in the Saglek Block, northern Labrador, Canada: the oldest geological evidence for plate tectonics. *Tectonophysics* 662:40–66
- Lange MA, Ahrens TJ (1982) The evolution of the impact-generated atmosphere. *Icarus* 51:96–120
- Lissauer JJ, Stevenson DJ (2007) Formation of giant planets. In: Reipurth B et al (eds) *Protostars and planets V*. Univ Arizona Press, Tucson, pp 591–606
- Marchi S, Bottke WF et al (2014) Widespread mixing and burial of Earth's Hadean crust by asteroid impacts. *Nature* 511:578–582
- Maruyama S, Komiya T (2011) The oldest pillow lavas, 3.8–3.7 Ga from the Isua supracrustal belt, SW Greenland: plate tectonics had already begun by 3.8 Ga. *J Geogr* 120:869–876
- Melosh HJ, Vickery AM (1989) Impact erosion of the primordial atmosphere of Mars. *Nature* 338:487–489
- Miller SL (1953) A production of amino acids under possible primitive earth conditions. *Science* 117:528–529
- Mojzsis SJ, Harrison TM et al (2001) Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* 409:178–181
- Morbidelli A, Chambers J et al (2000) Source regions and timescales for the delivery of water to the earth. *Meteorit Planet Sci* 35:1309–1320
- Natta A, Grinin V et al (2000) Properties and evolution of disks around pre-main-sequence stars of intermediate mass. In: Mannings V et al (eds) *Protostars and planets IV*. Univ Arizona Press, Tucson, pp 559–588
- Neukum G, Ivanov BA (1994) Crater size distribution and impact probabilities on earth from lunar, terrestrial-planet, and asteroid cratering data. In: Gehrels T (ed) *Hazards due to comets and asteroids*. Univ Arizona Press, Tucson, pp 359–416
- O'Brien DP, Walsh KJ et al (2014) Water delivery and giant impacts in the 'grand tack' scenario. *Icarus* 239:74–84
- O'Neil J, Carlson RW et al (2008) Neodymium-142 evidence for hadean mafic crust. *Science* 321:1828–1831
- Ozima M, Podosek FA (2002) *Noble gas geochemistry*, 2nd edn. Cambridge Univ Press, Cambridge
- Ramirez RM, Kopparapu R et al (2013) Warming early Mars with CO₂ and H₂. *Nat Geosci* 7:59–63
- Raymond SN, O'Brien DP et al (2009) Building the terrestrial planets: constrained accretion in the inner solar system. *Icarus* 203:644–662
- Sagan C, Mullen G (1972) Earth and Mars: evolution of atmospheres and surface temperatures. *Science* 177:52–56
- Schaefer L, Fegley B (2007) Outgassing of ordinary chondritic material and some of its implications for the chemistry of asteroids, planets, and satellites. *Icarus* 186:462–483
- Schaefer L, Fegley B (2010) Volatile element chemistry during metamorphism of ordinary chondritic material and some of its implications for the composition of asteroids. *Icarus* 205:483–496
- Schlesinger G, Miller SL (1983) Prebiotic synthesis in atmospheres containing CH₄, CO, and CO₂. I. Amino acids *J Mol Evol* 19:376–382

- Shuvalov V (2009) Atmospheric erosion induced by oblique impacts. *Meteorit Planet Sci* 44:1095–1105
- Svetsov VV (2007) Atmospheric erosion and replenishment induced by impacts of cosmic bodies upon the earth and Mars. *Sol Syst Res* 41:28–41
- Tajika E, Matsui T (1992) Evolution of terrestrial proto-CO₂ atmosphere coupled with thermal history of the earth. *Earth Planet Sci Lett* 113:251–266
- Tera F, Papanastassiou DA et al (1974) Isotopic evidence for a terminal lunar cataclysm. *Earth Planet Sci Lett* 22:1–21
- Trail D, Watson EB et al (2011) The oxidation state of Hadean magmas and implications for early Earth's atmosphere. *Nature* 480:79–82
- Tyburczy JA, Frisch B et al (1986) Shock-induced volatile loss from a carbonaceous chondrite: implications for planetary accretion. *Earth Planet Sci Lett* 80:201–207
- Urey HC (1952) On the early chemical history of the earth and the origin of life. *Proc Natl Acad Sci U S A* 38:351–363
- Walsh KJ, Morbidelli A et al (2011) A low mass for Mars from Jupiter's early gas-driven migration. *Nature* 475:206–209
- Warren PH (2011) Stable-isotopic anomalies and the accretionary assemblage of the Earth and Mars: a subordinate role for carbonaceous chondrites. *Earth Planet Sci Lett* 311:93–100
- Wilde SA, Valley JW et al (2001) Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* 409:175–178

Chapter 15

Biogenic and Abiogenic Graphite in Minerals and Rocks of the Early Earth



Takeshi Kakegawa

Abstract Minerals and rocks older than 3.7 billion years in age have attracted investigators looking to find evidence for traces of life. However, ancient rocks are heavily metamorphosed, leaving only graphitized organic matter. In addition, abiogenic graphite is ubiquitous in ancient rocks, creating difficulties for researchers seeking to claim true traces of life. Previous investigations searching for traces of life, based in Jack Hills (Australia), Akilia and Isua (Greenland), and Nuvvuagittuq (Canada), are summarized in this chapter. Graphite accompanied by low $\delta^{13}\text{C}$ values was found at all these localities. However, whether or not this graphite has a biogenic origin is the subject of debate. Rocks from the >3.7 Ga Isua supracrustal belt provide robust evidence of traces of life in the form of well-preserved sedimentation features of reduced carbon accompanied by low $\delta^{13}\text{C}$ values. The occurrence of graphite with low $\delta^{13}\text{C}$ values in ancient rocks cannot be used to indicate the presence of a microbial biosphere on the early Earth unless the age of the carbonaceous material is confirmed as the same age of host rock sedimentation.

Keywords Biogenic graphite · Archean · Early earth · Biosphere · STEM

15.1 Introduction

15.1.1 Graphite Formation in Metamorphic Rocks

Hadean is the age before ca. 4.0 Ga. Hadean minerals have been preserved at a few localities, but the rocks have not survived. Archean is the age between ca. 4.0 and 2.5 Ga. Early Archean rocks appear at several localities, specifically in the North American continent (Fig. 15.1). Previous researchers have searched for evidence of traces of life in these ancient minerals and rocks. The earliest life was most likely a single-cell microbial organism. The oldest cell-like fossils have been found in

T. Kakegawa (✉)

Graduate School of Science, Tohoku University Geosciences, Miyagi, Japan
e-mail: kakegawa@m.tohoku.ac.jp

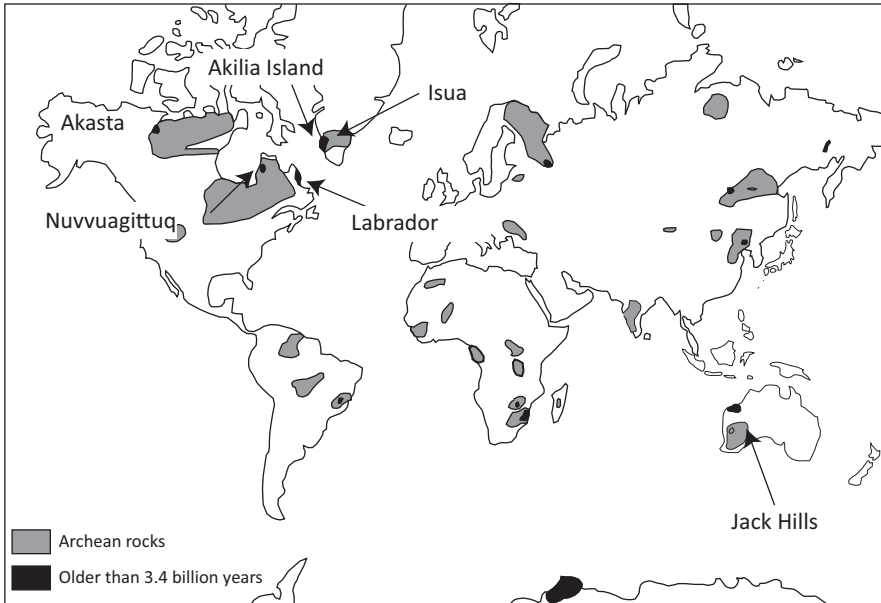


Fig. 15.1 Distribution of Hadean minerals (Jack Hills) and early Archean rocks. Hadean is the age older than ca. 4.0 Ga. There are no rock records from the Hadean, although minerals from this age are found

low-grade metamorphic rocks dated between the middle and late Archean (see Chap. 16, Sugitani ([this volume](#))). Rocks older than 3.5 Ga have been altered by high pressure (P) and temperature (T) metamorphism, and, therefore, the texture and chemistry of cells have been lost. Only a part of cells' components, i.e., thermally altered organic matter, survives in these metamorphosed rocks as graphite.

Abiogenic graphite can be formed by chemical processes under metamorphic temperature and pressure conditions (Fig. 15.2). Abiogenic graphite is generated either by (1) precipitation of graphite from metamorphic fluids by mixing CO_2 and CH_4 (Naraoka et al. 1996), (2) disproportionation of FeCO_3 (van Zuilen et al. 2002), or (3) Fischer-Tropsch-type reaction (McCollom and Seewald 2007, Fig. 15.2). Abiogenic graphite is ubiquitous in rocks and minerals formed on the early Earth. This creates a problem in distinguishing true traces of life in ancient rocks. In this article, studies on traces of life are summarized, and the identification of biogenic graphite in ancient rocks is discussed.

15.1.2 Characterization of Graphite

Studies of X-ray diffraction and high-resolution transmission electron microscopy (HRTEM) are frequently used to characterize graphite in ancient rocks (e.g., Grew 1974). However, they require extraction of graphite or organic matter from host

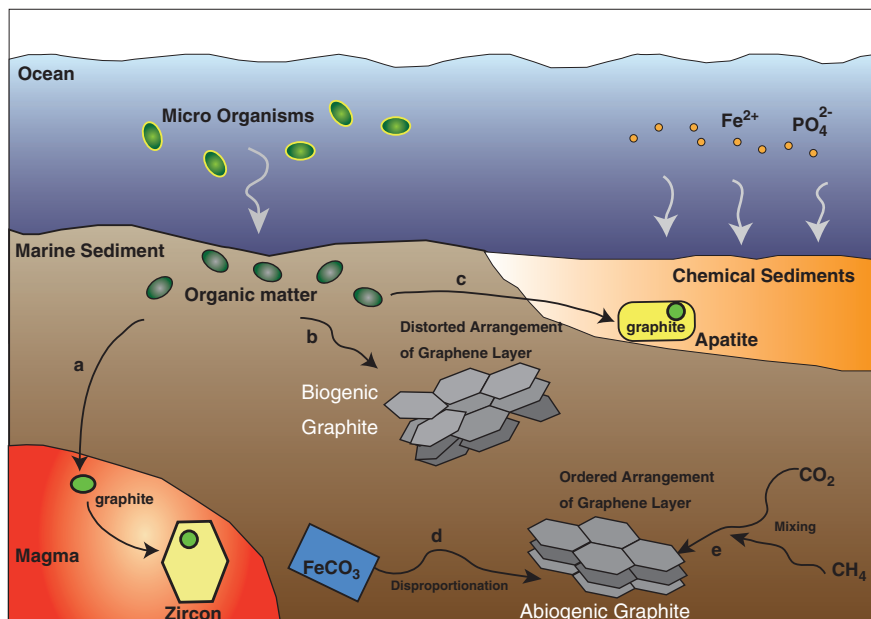


Fig. 15.2 Comparison of proposed models for origin of graphite in Hadean to early Archean metamorphosed rocks. (a) Granitic magma was generated by melting pre-existing sedimentary rocks which contained biogenic organic matter. During solidification of magma, zircon crystals were formed. Such zircon trapped thermally altered organic matter in crystals as impurities. (b) Biogenic organic matter in marine sediments were converted to graphite by metamorphism. Carbon atoms are bonded together forming hexagonal honeycomb lattice (i.e., graphene, gray-colored part in this figure). Layers of graphene stacked on top of each other form graphite. Biogenic graphite has distorted arrangement of graphene layers or disordered (irregular) stacking. (c) Apatite crystals were formed in chemical marine sediments such as banded iron formation. Biogenic organic matters were trapped in those apatite crystals when they were grown in sediments. (d) Fe-carbonate becomes unstable when temperature exceeds 550 °C and disproportionate into graphite and Fe-oxides. This chemical process produces typical abiogenic graphite. Abiogenic graphite shows well-ordered arrangement or very regulated stacking patterns of graphene layers. (e) Abiogenic graphite is also formed by mixing CO₂-rich and CH₄-rich fluids at metamorphic conditions

rocks. Concentrations of carbonaceous materials in metamorphosed rocks are not high (e.g., 0.1–0.4 wt % C; Rosing 1999), so treatment of large quantities of rock samples is often required to extract carbonaceous materials. Less metamorphosed organic matter shows a broad peak in X-ray diffraction patterns and so is not suitable for X-ray diffraction studies. The progression of graphitization of organic matter with metamorphic temperature yields the sharp peak of graphite in X-ray diffraction patterns. The X-ray diffraction patterns of graphite are used to evaluate metamorphic grade (Wada et al. 1994).

Extracted graphite or organic matter can be studied using high-resolution transmission electron microscopy (HRTEM, Buseck and Huang 1985). Graphite is com-

posed of multiple layers of graphene (Fig. 15.2). The carbon atom arrangement in graphite or the stacking patterns of graphene can be seen in HRTEM images. Less metamorphosed organic matter does not show clear HRTEM images.

An alternative analytical method is provided by laser Raman microspectroscopy (e.g., Wopenka and Pasteris 1993). Raman microspectroscopy is the most appropriate tool for quantitative characterization of graphitized organic matter. The Raman spectrum of graphitized organic matter is composed of first-order (1100–1800 cm^{-1}) and second-order (2500–3100 cm^{-1}) regions (Beysac et al. 2002a). The first-order region (Fig. 15.3), in which the graphite band (G band) occurs at 1580 cm^{-1} , is frequently used for metamorphic rocks. Other first-order bands appear around 1350 cm^{-1} and 1620 cm^{-1} . The band occurring at 1350 cm^{-1} (D1 band) is intense and very wide in poorly ordered carbons. This band has been attributed to the presence of heteroatoms (O, H, N) or structural defects. The 1620 cm^{-1} band (D2 band) appears as a shoulder on the G band. The 1500 cm^{-1} band (D3 band) is present in poorly ordered carbons as a very wide band. Arrowed peak in Fig. 15.3 is the raw data for apparent “G” band. This arrowed peak is made of combination of real G, D2, and a part of D3 bands. Computer calculations on this apparent “G” band (arrowed band) separate true G, D2, and D3 bands. The intensity of the G band relative to the D bands correlates with the degree of graphitization. When the carbon atom arrangement in graphite is disordered, high and broad D1 to D3 peaks are relatively clear. The degree of graphitization of organic matter is, in general, correlated with metamorphic temperature. Therefore, Raman spectra of graphitized organic matter are commonly used as geothermometers to estimate metamorphic temperatures (Beysac et al. 2002a, b).

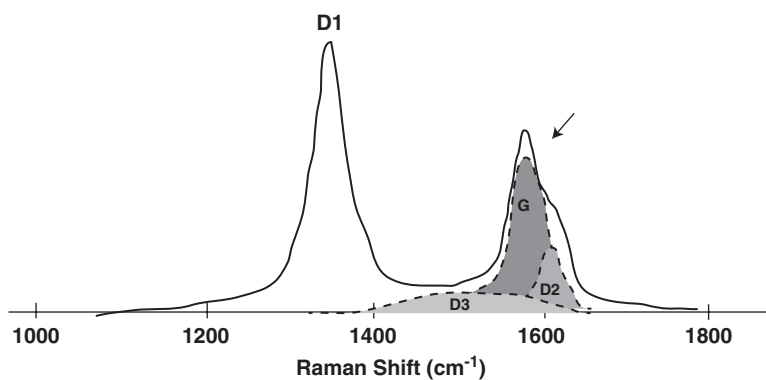


Fig. 15.3 Raman spectrum. The graphite band (G band) occurs at 1580 cm^{-1} . The band occurring at 1350 cm^{-1} (D1 band) is intense and very wide and found in poorly ordered carbons. This band has been attributed to the presence of heteroatoms (O, H, N) or structural defects. The 1620 cm^{-1} band (D2 band) appears as a shoulder on the G band. The 1500 cm^{-1} band (D3 band) is present in poorly ordered carbons as a very wide band. Arrowed peak in Fig. 15.3 is the raw data for apparent “G” band. This arrowed peak is made of combination of real G, D2, and a part of D3 bands

15.1.3 Traditional Tool to Identify Traces of Life: Carbon Isotope Compositions

Graphite has two stable isotopes: ^{12}C and ^{13}C . Carbon isotope compositions are traditionally used as a standard tool to distinguish biogenic graphite from abiogenic graphite. Carbon isotope compositions are expressed using δ -notation.

$$\delta^{13}\text{C} = \left(\left(\frac{^{13}\text{C}}{^{12}\text{C}} \right)_{\text{(sample)}} / \left(\frac{^{13}\text{C}}{^{12}\text{C}} \right)_{\text{(PDB)}} - 1 \right) \times 1000 \text{ (per mil)}$$

PDB is a Cretaceous *belemnite* fossil in the Pee Dee Formation and is used as an international standard for carbon isotope calibration. Gas mass spectrometry is used for conventional analysis of the $\delta^{13}\text{C}$ values of carbonaceous materials. For this analytical method, graphite and/or organic matter must be extracted from the host rocks. Extracted carbonaceous materials are then combusted into CO_2 , which is then introduced into a gas mass spectrometer to determine the $^{13}\text{C}/^{12}\text{C}_{\text{(sample)}}$ ratio.

Separation of graphite and organic matter is not necessary when $\delta^{13}\text{C}$ values are analyzed using secondary ionization mass spectrometry (SIMS). An accelerated and focused primary ion hits a spot on the targeted samples with an area of a few square nanometers. The targeted spot is then ionized, and some atoms are sputtered as secondary ions. The secondary ions are filtered according to atomic mass, and their isotope compositions are then measured. However, the accuracy of SIMS is less than that of the conventional gas mass spectrometry method. SIMS analyses often produce erroneous $\delta^{13}\text{C}$ values. This issue requires special attention in the evaluation or correction of carbon isotope data obtained by SIMS (House 2015).

$\delta^{13}\text{C}$ values for graphite and/or biogenic organic matter are determined either by equilibrium or kinetic isotope fractionation of carbon isotopes. Figure 15.4 shows $\delta^{13}\text{C}$ values for equilibrium isotope fractionation at various temperatures. The fractionation factor is expressed as $1000\ln\alpha$, which is the difference between $\delta^{13}\text{C}$ values of graphite and other carbon species. CO_2 , which $^{12}\text{C}/^{13}\text{C}$ ratio is known, is taken as a reference carbon species, and the isotope difference between the reference CO_2 and graphite is expressed as the following equation: $1000\ln\alpha = \delta^{13}\text{C}_{\text{(CO}_2\text{)}} - \delta^{13}\text{C}_{\text{(graphite)}}$. $1000\ln\alpha$ is a function of temperature. To simplify the discussion, it is assumed that graphite and CO_2 gases are isotopically equilibrated in a closed space. The $\delta^{13}\text{C}$ value of the initial CO_2 is also assumed to be -6‰ . Typical metamorphic temperature to form abiogenic graphite is greater than 500°C . According to Fig. 15.4, $\delta^{13}\text{C}$ value of graphite is -16‰ at 500°C . This could be the lowest $\delta^{13}\text{C}$ value for abiogenic graphite in this simplified system. Indeed, $\delta^{13}\text{C}$ values of many abiogenic graphite samples from natural rocks are less than -16‰ (van Zuilen et al. 2002).

Photosynthesizing or chemoautotrophic bacteria are common primary producers in ancient and modern aquatic environments. When these microorganisms use atmospheric CO_2 or dissolved CO_2 in water for metabolism, they produce organic matter depleted in ^{13}C . This is because of preferential partitioning of ^{12}C into organic matter by biological activities. This metabolism-induced fractionation is termed kinetic isotope fractionation associated with biological activities.

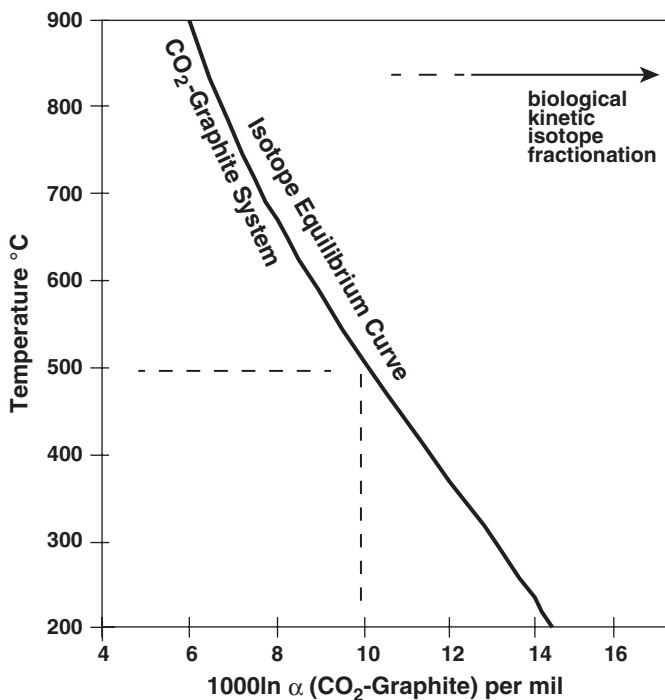


Fig. 15.4 Equilibrium isotope fractionation. (Data from Bottinga (1968))

Because of different isotope fractionation systems, $\delta^{13}\text{C}$ values for biogenic organic matter are different from the $\delta^{13}\text{C}$ values of abiogenic graphite (indicated as arrow's direction in Fig. 15.4). For example, $\delta^{13}\text{C}$ values of cultured cyanobacteria are -30 to -20‰ (Schidlowski, 2000), which is more ^{12}C -enriched than abiogenic graphite (-16‰) formed at 500 °C . When organic matter is converted into graphite by geological processes, $\delta^{13}\text{C}$ values of the organic matter are shifted toward heavier values. But biogenic ^{12}C -enriched compositions remain in such graphite. Based on this principle, carbon isotope compositions of graphite have been used to distinguish abiogenic graphite from biogenic graphite in past literatures.

15.2 Traces of Life in Hadean and Early Archean Ages

15.2.1 Graphite in the Jack Hills Zircon

The oldest Earth materials are zircon crystals found in Jack Hills in western Australia (Fig. 15.1). The age of the zircon crystals ranges from 4.4 to 3.9 Ga, and 4.1 Ga zircon crystals have submicron-sized graphite inclusions. Bell et al. (2015) determined $\delta^{13}\text{C}$ values of these graphite inclusions using SIMS: $\delta^{13}\text{C}_{(\text{graphite})}$ values were

around $-24 \pm 5\%$. Such value is within the range for biogenic graphite (Schidlowski 2000). Bell et al. (2015) also found slightly disordered arrangements of carbon atoms in graphite structures using Raman spectroscopic analyses. Such disordered carbon is a typical feature of biogenic carbon in modern sedimentary rocks. Based on the combination of carbon isotope compositions and the disordered carbon atoms in the graphite, Bell et al. (2015) claimed a biogenic origin for the graphite in 4.1 Ga zircon crystals.

Figure 15.2 describes a potential mechanism for the formation of graphite inclusions in Jack Hills zircon crystals. Hadean carbonaceous sedimentary rocks were melted in the deep crust, generating granitic magma. Zircons then crystallized as the granitic magma cooled. Accordingly, organic matter in the sedimentary rocks became graphite in the magma and was then incorporated into zircons.

In general, melting temperature of rocks exceeds $800\text{ }^{\circ}\text{C}$ (Fig. 15.2). This indicates that “biogenic graphite” should have experienced high temperatures, and the carbon atom arrangement of graphite becomes ordered at magmatic temperatures. However, Raman spectra of the graphite in Jack Hills zircons indicate a disordered carbon atom arrangement, which is puzzling if this “disordered” graphite represents >4.1 Ga biogenic organic matter that has been graphitized and survived in magma.

The abiotic origin of ^{13}C -depleted carbon isotope compositions of graphite in Jack Hills should also be considered. The fractionation mechanism producing isotopically light carbon in graphite may be an abiogenic process such as equilibrium fractionation involving CO_2 , CH_4 , and graphitic material (Horita 2001), Rayleigh distillation (Eiler 2007), surface-catalyzed processes (McCollom and Seewald 2006), or a combination of these processes. These factors suggest that the possibility of abiogenic graphite formation cannot be excluded when explaining the genesis of the graphite in Jack Hills zircons.

15.2.2 Akilia Island Debates

Akilia Island is a small island in western Greenland. The age of the rocks on Akilia Island is ca. 3.83 Ga (Nutman et al. 1997; Manning et al. 2006). Rocks with banded texture appear on the western edge of Akilia Island (Fig. 15.5). Mojzsis et al. (1996) considered that these rocks were Fe-rich chemical sediments deposited on the 3.83 Ga ocean floor. This banded rock contains microscopic apatite, which is a common phosphate mineral in marine sediments (Fig. 15.2). Mojzsis et al. (1996) reported that the apatite in the banded rocks encapsulates graphite inclusions with $\delta^{13}\text{C}$ values ranging from -49% to -21% . The carbon isotope composition of the graphite inclusions appeared consistent with metabolism-induced fractionation (Mojzsis et al. 1996; McKeegan et al. 2007). It was therefore concluded that this graphite represents early Archean metamorphosed bioorganic matter.



Fig. 15.5 Photograph of banded rocks from Akilia Island. A sample studied by Mojzsis et al. (1996) was collected from this outcrop (around the hammer in Figure)

Fedo and Whitehouse (2002) studied the same banded rocks. They found that the examined rocks are composed of quartz, amphibole, and pyroxene, which are unusual minerals in metamorphosed marine sediments. In addition, subsequent studies using the same sample have shown the presence of carbonaceous material at apatite surfaces (Nutman and Friend 2006) but have failed to identify graphite inclusions in the apatite crystals (Lepland et al. 2005; Nutman and Friend 2006). The lack of carbon detection inside the apatite is a serious problem and has raised questions about the original claim made by Mojzsis et al. (1996).

On the other hand, McKeegan et al. (2007) found a disordered arrangement of carbon atoms in the graphite in the same rocks using Raman spectroscopic analyses. The ^{13}C -depleted isotopic compositions of graphite support the interpretation that the graphite in the apatite may represent chemical fossils of early life (McKeegan et al. 2007). However, the carbon atom order of this graphite is lower than would be expected for a biogenic organic matter source that was converted into graphite under granulite facies conditions (Beysac et al. 2002b). Therefore, abiogenic processes, such as inorganic precipitation of graphitic material from carbonic fluids, can explain the observed occurrences of graphite in the Akilia samples (Lepland and Whitehouse 2011; Lepland et al. 2011). Though the possibility of biogenic graphite in the banded rocks at Akilia Island cannot be excluded, considerable uncertainty remains, and it is difficult to distinguish from the ubiquitous abiogenic graphite in the Akilia rocks, which suffered granulite facies metamorphism.

15.2.3 *Nuvvuagittuq*

The Nuvvuagittuq supracrustal belt (NSB) is a metamorphic volcano-sedimentary succession located in the Archean Superior craton along the eastern shore of Hudson Bay in northern Quebec (Fig. 15.1). The age of the NSB is at least 3770 million and possibly 4280 million years old (O'Neil et al. 2008). Dodd et al. (2017) found microfossil-like materials in metamorphosed ferruginous sedimentary rocks from the NSB. These structures occur as micron-sized tubes and filaments made of hematite crystals (Fe_2O_3). Similar tube and filament structures are found in modern submarine hydrothermal discharge areas. These modern structures represent fossilized sheaths of Fe-oxidizing bacteria. Therefore, Dodd et al. (2017) proposed that the NSB's Fe-oxide tubes and filaments are products of microorganisms. The characteristics of Fe-rich metasediments suggest that the above microorganisms most likely lived around submarine hydrothermal vents. In addition, isotopically light carbonaceous materials and carbonate minerals are found in these rocks, providing additional evidence for biological activity more than 3770 million years ago.

In contrast, Papineau et al. (2011) found poorly crystalline graphite with a disordered carbon atom arrangement in banded iron formations of the NBS using Raman spectroscopy. Average $\delta^{13}\text{C}_{(\text{graphite})}$ values were $-22.8 \pm 1.9\%$, implying a biological source for this carbon. However, the graphite experienced much lower temperatures than the host rocks during metamorphism. Therefore, the poorly crystalline graphite in these rocks was generated after peak metamorphism. This suggests that the graphite was not syngenetic with host rock. This invites skepticism as to whether the Fe-oxide tubes and filaments found by Dodd et al. (2017) were truly biological products older than 3.77 Ga.

15.3 Graphite in the Isua Supracrustal Belt (ISB)

The ISB in southwestern Greenland is part of the Itsaq Gneiss Complex. Rocks of the ISB consist predominantly of early Archean metavolcanics and minor metasedimentary rocks (e.g., Nutman et al. 2009) (Fig. 15.1). U-Pb zircon dating indicates that southern and northern Isua rocks show contrasting ages. It is thought that both the ca. 3.8 Ga southern and the ca. 3.7 Ga northern juvenile crustal complexes were tectonically juxtaposed along mylonite zones during the early Archean (Nutman et al. 2009, 2013). Most of the ISB was strongly deformed by early Archean tectonism (Nutman et al. 2015) and converted into amphibolite metamorphic facies, but not to granulite facies (Boak and Dymek 1982). The metamorphosed basaltic rocks, termed "garbenschiefer," occupy the central portion of the northwestern ISB with relict pillow lava structures.

Rosing (1999) reported turbiditic metasedimentary rocks that are >3.7 billion years old (Ga) in the garbenschiefer unit (Fig. 15.6). Modern turbidites are deposited in the deep ocean by underwater avalanches that move down slopes of the continental



Fig. 15.6 Photograph of metamorphosed black shale (schist) discovered by Rosing (1999)

shelf. When avalanched materials approach the bottom of the deep ocean, sands are deposited first followed by finer clays. This sequential sedimentation leaves a combination of sandstone-shale deposits. The rock reported by Rosing (1999) consists of both metamorphosed shale and sandstone (Fig. 15.6). This finding is clear evidence of early Archean marine sediments. The clear sedimentation and less metamorphosed features are different from the Akilia rocks, which could be magmatic in origin. The metamorphosed shale of the ISB has graphite contents of up to 0.4 wt % C. Carbon isotopic compositions of the graphite range from -19.1 to -18.7‰ (Rosing 1999). These isotope compositions are similar to the $\delta^{13}\text{C}$ values of organic matter in modern marine sedimentary rocks. Therefore, Rosing (1999) claimed that vast microbial ecosystems were present in early Archean oceans.

The metamorphosed shale often contains garnet crystals, which contain carbonaceous inclusions that are contiguous with carbon-rich shale beds. Hassenkam et al. (2017) investigated these carbonaceous inclusions using the in situ infrared absorption method. The absorption spectra are most likely to represent carbon bonded to nitrogen and oxygen and probably also to phosphate. These findings were the first to report detection of oxygen-, nitrogen-, and phosphate-related biological activities in rock records older than 3.7 Ga.

Nutman et al. (2016) reported 3700-Myr-old metacarbonate rocks in the ISB. The chemical compositions of the metacarbonates indicate shallow seawater depositional environments. Stromatolites, a few cm high, are found in such metacarbonates. If ISB stromatolites represent a true microbial community from 3.7 Ga shallow oceans, this indicates that photosynthesizing bacteria were already present, facilitat-

ing CO₂ sequestration. On the other hand, ISB stromatolites do not contain carbonaceous materials (Nutman et al. 2016). This fact remains confusing if ISB stromatolites are truly biogenic in origin, and an alternative model has been proposed for the origin of Isua's stromatolite by different researchers (Allwood et al. 2018).

15.4 Discovery of New Graphitic Schist in Isua

15.4.1 Occurrence of Graphitic Schist

Metamorphosed shale, which has become schist, was found on the northwestern side of the ISB by the author's research group (Fig. 15.7; Ohtomo et al. 2014). The locality lies in the garbenschiefer unit in the western part of the ISB, which is about 1 km west of the location of the turbiditic rocks (Rosing 1999). Three layers of BIFs are concordantly intercalated with metabasalts. The schist preserves sedimentary textures with various detrital minerals. The chemical compositions of the schist support the hypothesis that the precursor of the schist was shale. This result made it clear that clastic marine sediments are a robust finding in the ISB.

The schist contains abundant graphite (0.1–8.8 wt% C). Graphite occurs according to the compositional layering of the schist, which strikes parallel to that of the surrounding BIFs. The geothermometric signals of the Raman spectra for the schist indicate that the peak metamorphic temperatures of the graphite were $525 \pm$

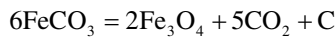


Fig. 15.7 Photograph of metamorphosed black shale (schist) discovered by Ohtomo et al. (2014)

50 °C. Prograde metamorphic temperatures for the ISB were reported as 500–600 °C (Boak and Dymek 1982), consistent with the Raman geothermometer result for graphite. This temperature consistency suggests that the precursor of the graphite was already present in the host rocks before prograde metamorphism. All features of the schist suggest a sedimentary origin for the reduced carbon and the syngeneity of the carbonaceous material in the host rock.

15.4.2 Carbon Isotope Compositions

Graphite occurs in various rocks in wide area of the ISB (Schidlowski et al. 1979; Ueno et al. 2002). It has been proposed that the majority of the graphite in the ISB was formed by inorganic chemical reactions under high temperature and pressure conditions (van Zuilen et al. 2002). For example, a carbonate vein, located in the northeastern part of the ISB, contains high concentrations of graphite (Fig. 15.8; 4.1 wt% C). Disproportionation of Fe-carbonate was suggested to explain the genesis of this graphite in carbonate veins. In these veins, substantial amounts of magnetite and residual Fe-carbonate occur alongside the graphite.



The Raman spectrum of the graphite in the veins is nearly identical to that of the graphite in the schist. The estimated metamorphic temperatures for the graphite



Fig. 15.8 Photograph of a carbonate vein studied by van Zuilen et al. (2002)

(496 ± 50 °C) are close to peak metamorphic temperatures, suggesting a metamorphic origin for the graphite in the veins.

The $\delta^{13}\text{C}_{(\text{graphite})}$ values of the graphite in the schist range from -23.8% to -12.5% (average: -17.9%). These isotope variations do not represent original compositions because early graphite has exchanged isotopes with carbonic fluids during metamorphism. In general, metamorphic fluids contain CO_2 with carbon isotope compositions similar to mantle values (-5 to -6%). CO_2 in metamorphic fluids exchange carbon isotope compositions when the graphite in the schist is exposed to CO_2 . As a result, $\delta^{13}\text{C}_{(\text{graphite})}$ values were modified and achieved isotope equilibrium compositions between $\delta^{13}\text{C}_{(\text{graphite})}$ and $\delta^{13}\text{C}_{(\text{CO}_2)}$ values (c.f., Fig. 15.3). Graphite aggregates in the schist often show zoning in $\delta^{13}\text{C}_{(\text{graphite})}$ within a few mm. The $\delta^{13}\text{C}_{(\text{graphite})}$ values change from -23% to -14% (Ohtomo et al. 2014). Such zoning is unidirectional and more enriched in ^{13}C in outer parts of the graphite aggregates. This isotope feature represents carbon isotope exchange between graphite and metamorphic CO_2 . In other words, the lowest $\delta^{13}\text{C}_{(\text{graphite})}$ value of -23% is close to the original $\delta^{13}\text{C}_{(\text{graphite})}$ value and is within the range for organic matter in modern marine sediments.

Ohtomo et al. (2014) analyzed ^{13}C values of the abiogenic graphite, which formed by disproportionation of FeCO_3 . The ^{13}C value was $+10.5\%$, and this value is explained by equilibrium isotope fractionation among $\text{FeCO}_3\text{-CO}_2\text{-C}$ at metamorphic temperatures in the ISB. Therefore, clear differences in $\delta^{13}\text{C}$ values are present between graphite in carbonate veins and graphite in the schist (meta-black shale) in the ISB. Apparently such isotope differences indicate that ^{13}C -depleted graphite is biogenic in origin. However, there exists ambiguity to discuss the origin of graphite only using carbon isotope compositions. Therefore, additional evidence is desired to claim biogenic origin.

15.4.3 TEM and STEM (Scanning Transmitted Electron Microscopy) Observations on Graphite

HRTEM and STEM observations of graphite may provide useful information to constrain the origin of graphite. STEM images revealed that the morphology of nanoscale graphite grains is different in carbonate veins and schist (Fig. 15.9). Nanoscale graphitic polygonal grains, tube-like structures, and other morphologies are found in the schist (Fig. 15.9A). On the other hand, sheeted flake is the single morphology of the graphite in carbonate veins (Fig. 15.9B). HRTEM images also show contrasting features between two types of rocks (Fig. 15.10). The sheeted flakes in the carbonate vein show well-layered structures overall, which indicates well-ordered stacking of graphene layers (Fig. 15.10b). In contrast, curled structures are present in the inner portions of graphite grains in the schist (Fig. 15.10a). Some lattice fringes show distortion at surfaces and inside the graphite grains (arrow in Fig. 15.10a).

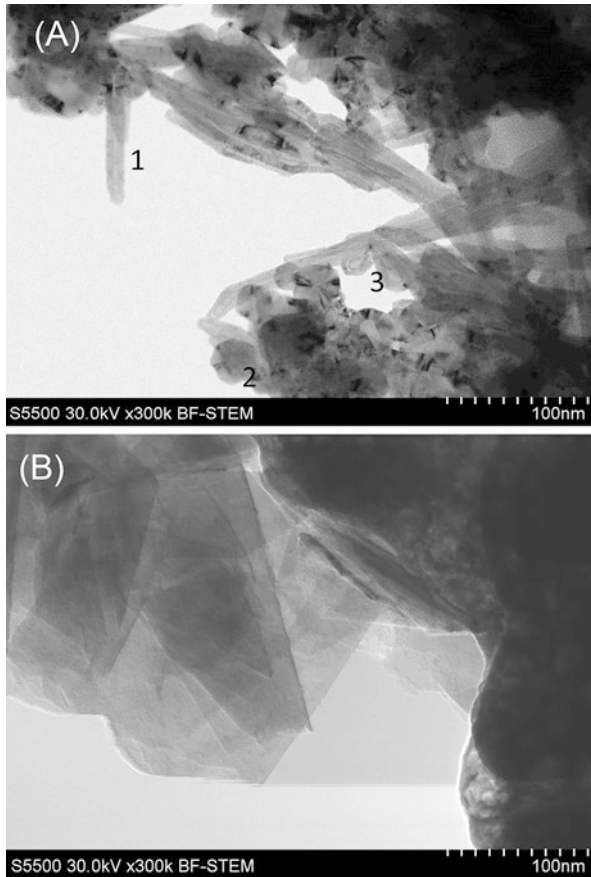


Fig. 15.9 Contrasting morphologies of graphite grains. A is from schist (meta-black shale) and B is from a carbonate vein. A shows various forms of graphite grains including tubular (1 in figure), polygonal (2 in figure), or intermediate (3 in figure) forms. Such variation is a typical characteristic of biogenic graphite. B shows uniform sheet-like textures. Individual thin sheets (or films) are aggregated. Such uniformity is a characteristic of abiogenic graphite

Organic matter contains non-graphitizing carbon, such as non-planar carbon ring compounds associated with abundant pores (Buseck and Huan 1985). Graphitization of organic matter becomes heterogeneous at nanoscale if source materials contain non-graphitizing carbon. Therefore, biogenic organic matter, which contains various molecules and functional groups, is favored as the precursor of the graphite in the schist. Various morphologies of graphite and distorted structure in graphene layers can be considered as heritage of biogenic organic matter. On the other hand, graphite in carbonate veins (abiogenic in origin) has uniform morphology of graphite and no distorted structures in graphene layers. Such contrast can be used to distinguish biogenic graphite from abiogenic graphite.

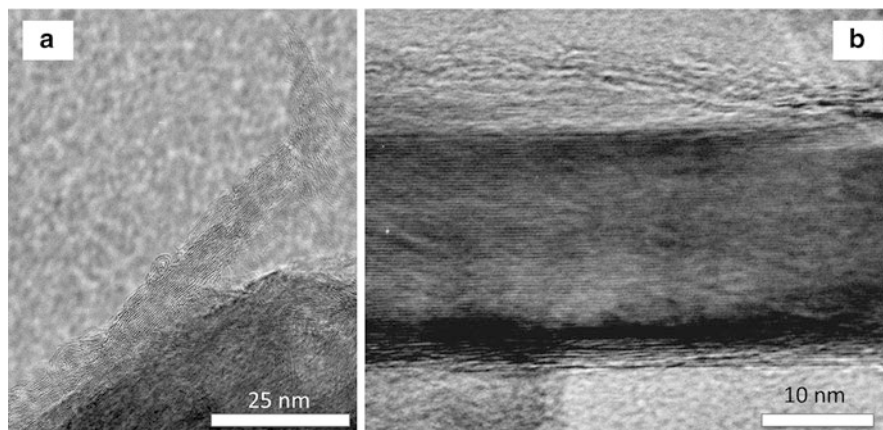


Fig. 15.10 HRTEM images of graphite from schist (**a**) and a carbonate vein (**b**) in the ISB. Thin lines in figures are individual graphene layers. Panel A shows rolled or rounded structure of graphene layers (arrowed part). Those textures are very common in and outside of graphite grains. Such rolled or rounded structure is a typical characteristic of biogenic graphite. Panel B shows well-stacked or well-arranged layers of graphene. Rounded structure is not found inside of graphite grains. Such well-stacked layer is a characteristic of abiogenic graphite

The combined information on geological occurrences, graphite morphologies, nanoscale structures, and isotopic compositions of the graphite in the schist suggest a biogenic origin. High concentrations of ^{13}C -depleted graphite in these rocks would require widespread biological activity to support the high rate of production and sedimentary delivery of organic matter to the >3.7 Ga ocean floor.

15.4.4 More Biogenic Carbon in Sedimentary Minerals

Mishima et al. (2016) found tourmaline-rich schist in the ISB. This schist is considered analogous to rocks examined by Ohtomo et al. (2014). Field relationships and chemical compositions suggest a sedimentary origin for the tourmaline-rich rocks. Garnet crystals contain a number of tourmaline inclusions (a type of borosilicate mineral). Both garnet and tourmaline often contain nanoscale inclusions of graphite (Fig. 15.11). Clay minerals in modern sediments have the capability to adsorb and concentrate borate, which could lead to boron enrichment during diagenesis, followed by tourmaline formation under metamorphic conditions (Furukawa and Kakegawa 2017). To have graphite inclusions in tourmaline, clay-rich sediments need to contain abundant organic matter. This sedimentary organic matter was converted into graphite during prograde metamorphism, while tourmaline was formed from borate-clay complexes. Therefore, graphite inclusions in tourmaline are also highly likely biogenic in origin.

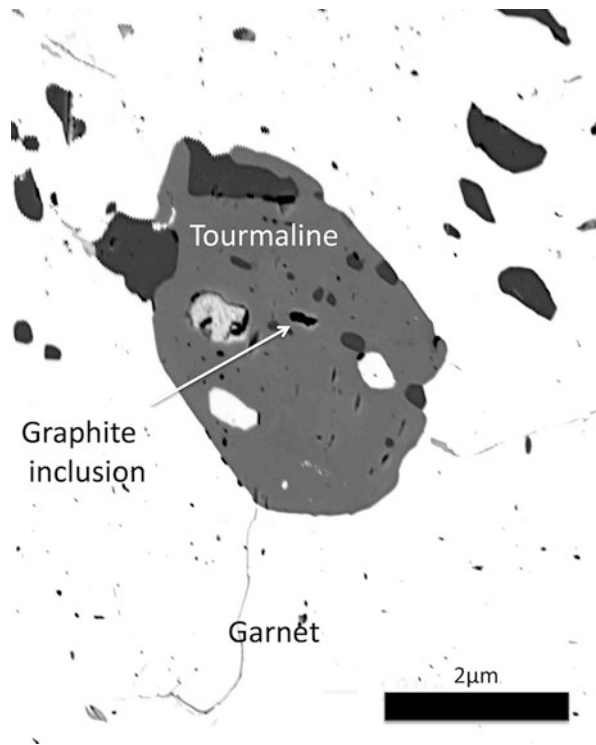


Fig. 15.11 Back-scattered electron image of a graphite inclusion in tourmaline. Scanning electron microscope (Hitachi S3000) was used to obtain this image. The host rock of the tourmaline is garnet-biotite schist, which occurs approximately 2 km south of the rocks studied by Ohtomo et al. (2014)

15.5 Conclusion

Hadean to Archean graphite crystals were found in various rocks and minerals from Jack Hills (Australia), Akilia and Isua (Greenland), and Nuvvuagittuq (Canada). Those graphite have ^{13}C -depleted carbon isotope compositions (low $\delta^{13}\text{C}$ values), apparently implying biogenic origin of graphite. However, some of those graphite are unlikely biogenic in origin, suggesting weakness of carbon isotope compositions to constrain origin of graphite.

Rocks from the >3.7 Ga Isua supracrustal belt contain graphite. Metamorphosed sedimentary rocks contain graphite. The age of graphite is confirmed as the same age of host rock sedimentation. Such graphite has unique nano-texture in individual graphite crystal accompanied with ^{13}C -depleted carbon isotope compositions. These analyses support the biological origin of the graphite in >3.7 Ga Isua supracrustal belt sedimentary rock. Studies of Isua rocks demonstrate how important to provide comprehensive information or to accumulate supportive evidence for traces of life in ancient rocks.

In the future, more evidence of biogenic graphite will probably be found in ancient rocks and minerals by using more advanced analytical techniques. On the other hand, it is necessary to examine the syngeneity of reduced carbon with host rocks more carefully when examining evidence for traces of life.

Acknowledgment This study was supported by JSPS Grants (#15H02144, 24403013).

References

- Allwood AC, Rosing MT, Flannery DT, Hurowitz JA, Heirweh CM (2018) Reassessing evidence of life in 3,700-million-year-old rocks of Greenland, *Nature* 563:241–245
- Bell EA et al (2015) Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. *Proc Natl Acad Sci USA* 112:14518–14521
- Beysac O, Goffe B, Chopin C, Rouzaud J-N (2002a) Raman spectra of carbonaceous material in metasediments: a new geothermometer. *J Metamorph Geol* 20:859–871
- Beysac O, Rouzaud JN, Goffé B, Brunet F, Chopin C (2002b) Graphitization in a high-pressure, low-temperature metamorphic gradient: a Raman microspectroscopy and HRTEM study. *Contrib Mineral Petrol* 143(1):19–31
- Boak JL, Dymek RF (1982) Metamorphism of the ca. 3800 Ma supracrustal rocks at Isua, west Greenland: implications for early Archaean crustal evolution. *Earth Planet Sci Lett* 59(1):155–176
- Bottinga (1968) Isotope fractionation in the system, Calcite-graphite-carbon dioxide-methane-hydrogen-water. California University at San Diego, Ph.D. Thesis, 126p
- Buseck PR, Huang B-J (1985) Conversion of carbonaceous material to graphite during metamorphism. *Geochim Cosmochim Acta* 49:2003–2016
- Dodd MS, Papineau D, Grenne T, Slack JF, Rittner M, Pirajno F, O’Neil J, Little CTS (2017) Evidence for early life in Earth’s oldest hydrothermal vent precipitates. *Nature* 543:60–64
- Eiler JM (2007) The oldest fossil or just another rock? *Science* 317:1046–1047
- Fedo CM, Whitehouse MJ (2002) Metasomatic origin of quartz-pyroxene rock, Akilia, Greenland, and implications for earth’s earliest life. *Science* 296:1448–1451
- Furukawa Y, Kakegawa T (2017) Borate and the origin of RNA: a model for the precursors to life. *Elements* 13:261–265
- Grew ES (1974) Carbonaceous material in some metamorphic rocks of New England and other areas. *J Geol* 82:50–73
- Hassenkam T, Anderson MP, Dalby KM, Mackenzie DMA, Rosing MT (2017) Elements of Eoarchean life trapped in mineral inclusions. *Nature* 548:78–81
- Horita J (2001) Carbon isotope exchange in the system CO₂-CH₄ at elevated temperatures. *Geochim Cosmochim Acta* 218:171–186
- House C (2015) A synthetic standard for the analysis of carbon isotopes of carbon in silicates, and the observation of a significant water-associated matrix effect. *Geochem Trans* 16:14. <https://doi.org/10.1186/s12932-015-0029-x>
- Lepland A, Whitehouse MJ (2011) Metamorphic alteration, mineral paragenesis and geochemical re-equilibration of early Archean quartz-amphibole-pyroxene gneiss from Akilia, Southwest Greenland. *Int J Earth Sci* 100:1–22
- Lepland A et al (2005) Questioning the evidence for Earth’s earliest life – Akilia revisited. *Geology* 33:77–79
- Lepland A, Van Zuilen MA, Philippot P (2011) Fluid-deposited graphite and its geobiological implications in early Archean gneiss from Akilia. *Greenland Geobiol* 9:2–9

- Manning CE, Mojzsis SJ, Harrison TM (2006) Geology, age and origin of supracrustal rocks at Akilia, West Greenland. *Am J Sci* 306:303–366
- McCollom TM, Seewald JS (2006) Carbon isotope composition of organic compounds produced by abiotic synthesis under hydro-thermal conditions. *Earth Planet Sci Lett* 243:74–84
- McCollom TM, Seewald JS (2007) Abiotic synthesis of organic compounds in deep-sea hydrothermal environments. *Chem Rev* 107:382–401
- McKeegan KD, Kudryavtsev AB, Schopf JW (2007) Raman and ion microscopic imagery of graphitic inclusions in apatite from older than 3830 Ma Akilia supracrustal rocks, West Greenland. *Geology* 35:591–594
- Mishima S, Ohtomo Y, Kakegawa T (2016) Occurrence of tourmaline in metasedimentary rocks of the Isua supracrustal belt, Greenland: implications for ribose stabilization in Hadean marine sediments. *Orig Life Evol Biosph* 46:247–271
- Mojzsis SJ, Arrhenius G, McKeegan KD, Harrison TM, Nutman AP, Friend CRL (1996) Evidence for life on Earth before 3,800 million years ago. *Nature* 384:55–59
- Naraoka H, Ohtake M, Maruyama S, Ohmoto H (1996) Non-biogenic graphite in 3.8-Ga metamorphic rocks from the Isua district, Greenland. *Chem Geol* 133:251–260
- Nutman AP, Friend CRL (2006) Petrography and geochemistry of apatites in banded iron formation, Akilia, W. Greenland: consequences for oldest life evidence. *Precambrian Res* 147:100–106
- Nutman AP, Mojzsis S, Friend CRL (1997) Recognition of 3850 Ma water-lain sediments in West Greenland and their significance for the early Archaean Earth. *Geochim Cosmochim Acta* 61:2475–2484
- Nutman AP, Friend CR, Paxton S (2009) Detrital zircon sedimentary provenance ages for the Eoarchaean Isua supracrustal belt southern West Greenland: juxtaposition of an imbricated ca. 3700 Ma juvenile arc against an older complex with 3920–3760 Ma components. *Precambrian Res* 172(3):212–233
- Nutman AP, Bennett VC, Friend CR, Hidaka H, Yi K, Lee SR, Kamiichi T (2013) The Itsaq Gneiss complex of Greenland: episodic 3900 to 3660 Ma juvenile crust formation and recycling in the 3660 to 3600 Ma Isukasian orogeny. *Am J Sci* 313(9):877–911
- Nutman AP, Bennett VC, Friend CR (2015) The emergence of the Eoarchaean proto-arc: evolution of a c. 3700 Ma convergent plate boundary at Isua, southern West Greenland. *Geol Soc Lond Spec Publ* 389(1):113–133
- Nutman AP, Bennett VC, Friend CR, Kranendonk MV, Chivas AR (2016) Rapid emergence of life shown by discovery of 3700-million-year-old microbial structures. *Nature* 537:535–538
- O’Neil J, Carlson RW, Francis D, Stevenson RK (2008) Neodymium-142 evidence for Hadean mafic crust. *Science* 321:1828–1831
- Ohtomo Y, Kakegawa T, Ishida A, Nagase T, Rosing MT (2014) Evidence for biogenic graphite in early Archaean Isua metasedimentary rocks. *Nat Geosci* 7:25–28
- Papineau D, Gregorio BT, Cofy GD, O’Neil J, Steele A, Stroud RM, Fogel ML (2011) Young poorly crystalline graphite in the >3.8-Gyr-old Nuvvuagittuq banded iron formation. *Nat Geosci* 4:376–379
- Rosing MT (1999) ^{13}C -depleted carbon microparticles in >3700-Ma sea-floor sedimentary rocks from West Greenland. *Science* 283(5402):674–676
- Schidlowski M (2000) Carbon isotopes and microbial sediments. In: Riding RE, Awramik SM (eds) *Microbial sediments*. Springer-Verlag, Berlin, pp 84–95
- Schidlowski M, Appel WU, Eichmann R, Junge CE (1979) Carbon isotope geochemistry of the 3.7×10^9 -yr-old Isua sediments, West Greenland: implications for the Archean carbon and oxygen cycles. *Geochim Cosmochim Acta* 43:189–199

- Sugitani K (this volume) Chapter 16: Cellular microfossils and possible microfossils in the Paleo and Mesoarchean. In: Yamagishi A, Kakegawa T, Usui T (eds) *Astrobiology*. Springer, Singapore
- Ueno Y, Yurimoto H, Yoshioka H, Komiya T, Maruyama S (2002) Ion microprobe analysis of graphite from ca. 3.8 Ga metasediments, Isua supracrustal belt, West Greenland: relationship between metamorphism and carbon isotope composition. *Geochim Cosmochim Acta* 66:1257–1268
- Van Zuilen MA, Lepland A, Arrhenius G (2002) Reassessing the evidence for the earliest traces of life. *Nature* 418:627–630
- Wada H, Tomita T, Matsuura K, Iuchi K, Ito M, Morikiyo T (1994) Graphitization of carbonaceous matter during metamorphism with references to carbonate and pelitic rocks of contact and regional metamorphisms, Japans. *Contrib Mineral Petrol* 118:217–228
- Wopenka B, Pasteris JD (1993) Structural characterization of kerogens to granulite-facies graphite: applicability of Raman microprobe spectroscopy. *Am Mineral* 78:533–557

Chapter 16

Cellular Microfossils and Possible Microfossils in the Paleo- and Mesoarchean



Kenichiro Sugitani

Abstract Representative Paleo- and Mesoarchean (>3.0 Ga) microfossils and possible microfossils retaining cellular structures from the Pilbara Craton, Western Australia, and the Kaapvaal Craton, South Africa, are reviewed. Rod-shaped, spheroidal, lenticular, and filamentous (and their subtypes) microfossils have been identified in those areas, and their sizes range from submicrons to 300 μm across. Depositional environments of host rocks vary from shallow marine or even terrestrial to deep-sea, with or without hydrothermal activities, providing no constraints on the geologic setting for the emergence of life. Although biological affinities such as cyanobacteria and sulfur bacteria have been proposed for a few types of Paleo- and Mesoarchean microfossils, those of most others are poorly understood.

Significantly, recent progress in Archean geobiology has revealed that the fossil record includes large (from 20 μm up to 300 μm along the major dimension), organic-walled spheroid and lenticular microfossils. If their biological affinities can be determined convincingly, they would provide us with new insights into the early biosphere and its evolution on the Earth and potentially on other planets. Further challenging and innovative studies are required in order to reveal the diversity of Paleo- and Mesoarchean ecosystems and to develop a taxonomy for such ancient microfossils.

Keywords Archean · Biotic diversity · Biological affinity · Large microfossils · Ecosystem

K. Sugitani (✉)

Graduate School of Environmental Studies, Nagoya University, Nagoya, Japan
e-mail: sugi@info.human.nagoya-u.ac.jp

16.1 Introduction

Evidence for Precambrian life has been accumulating since the discovery of the 1.9 Ga Gunflint microbiota more than 50 years ago (Barghoorn and Tyler 1965; Cloud 1965). It includes a wide range of categories, such as microfossils retaining cellular structures, microbial sedimentary structures, biomolecules, bio-minerals, and isotopic signatures, providing us with insights into the early evolution of life and its relevant ecosystems. In particular, cellular microfossils give us direct illustrations of ancient organisms and thus provide the opportunity to compare them with younger and extant organisms. Although Archean paleontology has sometimes been subjected to severe criticism and skepticism (Brasier et al. 2002, 2005, 2006; Lindsay et al. 2005; Wacey et al. 2018a, b), new discoveries of cellular microfossils have subsequently been published (e.g., Sugitani et al. 2007, 2010; Wacey et al. 2011, Javaux et al. 2010; Homann et al. 2016; Kremer and Kaźmierczak 2017; Schopf et al. 2017). It now seems widely accepted that complex ecosystems composed of diverse microbes had already evolved by 3.0 Ga, although newly described potential biosignatures, such as a 3.7 Ga “stromatolite” from Greenland (Nutman et al. 2016); >3.8 Ga hematitic filaments from the Nuvvuagittuq belt, Canada (Dodd et al. 2017); and a ca. 3.5 Ga microbial palisade fabric associated with geyserite (Djokic et al. 2017), are needed to be confirmed by follow-up studies (e.g., Witze 2016). Although some informative reviews have been published (Schopf and Walter 1983; Alterman and Kaźmierczak 2003; Schopf 2006; Wacey 2009, 2012), it seems worthwhile reviewing the Paleo- and Mesoarchean microfossil record more extensively, considering the recent progress in Archean paleobiology. In this chapter, I provide a detailed review of representative cellular microfossils and possible microfossils from pre-3.0 Ga successions (including assumed maximum ages) from South Africa and Western Australia, emphasizing the very early diversification of microorganisms and their adaptations to various environments, which provide a framework for future studies.

16.2 Criteria for the Biogenicity of Microbe-Like Structures

It is absolutely essential in any studies of Archean paleobiology to demonstrate the biogenicity of microbe-like structures. The criteria for assessing microbial structures in Archean rocks are therefore discussed before description of selected microfossils. The criteria are here consolidated and simplified from those previously proposed by Schopf and Walter (1983), Buick (1990), Brasier et al. (2002, 2005, 2006), Hofmann (2004), Sugitani et al. (2007), and Wacey (2009) and are shown in Table 16.1. Brasier et al. (2002, 2005, 2006) and Wacey (2009) claimed that Archaean microfossil-like structures needed to be considered as non-biological until the non-biological origin of the structures is refuted. However, this “null hypothesis” approach is too exclusive because it is virtually impossible to test “all

Table 16.1 Criteria for biogenicity of Archean microbe-like structures

<i>Geological context</i>
Host rocks of microstructures should be in well-known Archean terranes and certainly be a part of a geographically extensive Archean sedimentary succession. Preferably, the age of the host rock should be determined directly; alternatively, an established age for the unit from which rock specimens were collected could be substituted. Veins cannot be precluded as research targets, but the timing of their formation must be clearly elucidated. Lower metamorphic grade of host rocks are generally considered more suitable for the preservation of microfossils
<i>Indigenosity and syngenicity</i>
Indigenosity means that the microbe-like structures are demonstrably present in the host rocks. Structures in petrographic thin sections are usually indigenous, provided that they are in primary facies. Syngenicity requires structures to have been embedded during the primary phase of sedimentation or precipitation. In other words, the age of the embedding of the structure should be identical to the age of the deposition of host rocks. If this is verified, structures present within pores, fractures, and veins could be legitimate research targets
<i>Biological context</i>
<i>Size</i>
Unicellular organisms have a wide size range (from ~200 nm for <i>Mycoplasma</i> up to ~20 cm for coenocytic <i>Xenophyophore</i>); consequently size should not be a criterion for biogenicity. Rather, the size range is more important; populations of microbe-like structures would be expected to have a narrow size range
<i>Shape</i>
Cellular elaborations, such as organelles, flagella, and appendages, are key criteria for biogenicity, although these delicate structures are rarely preserved. Regardless of their origins and functions, the more elaborate the structures, the more certain their biogenicity
<i>Occurrence</i>
Microbe-like structures need to be abundant. Presence in colony-like clusters, including composite structures potentially resulting from reproduction, is a good indicator of biogenicity
<i>Taphonomy</i>
Organic compounds that form cell envelopes have plasticity and are subject to postmortem degradation that increases with diagenesis. Genuine microfossils show a range of wall preservation ranging from hyaline to granular, and taphonomic features, such as folding, breakage, tearing, and corrosion, can be identified
<i>Chemical and isotopic composition</i>
A carbonaceous composition (commonly as kerogen) is a key geochemical criterion for the biogenicity, although organic matter can also be mobilized and redistributed in the rock matrix. An association of S, N, and hopefully P within the carbonaceous structures provides positive evidence of biogenicity. A light (< -20 per mil) carbon isotopic value is a widely accepted biogenicity criterion, although it is not always diagnostic

possible alternatives” for biological origins of microfossil-like objects. A better approach is to employ the descending scale of probability referred to by Schopf et al. (1983) and Awramik and Grey (2005):

Compelling evidence: abundant evidence that permits only one reasonable interpretation

Presumptive evidence: the preponderance of evidence that suggests a most likely interpretation, but less probable interpretations also merit consideration.

Permissive evidence: evidence that seems consistent with at least two more or less equally tenable competing interpretations.

Suggestive evidence: evidence that although weak is at least consistent with the interpretation.

Missing evidence: there is no direct interpretable evidence to support the interpretation and to employ prefixes such as “pseudo-” and “dubio-” and the adjectives “putative-,” “possible-,” “probable-,” and “genuine-” in order to express the reliability of the biological origin of the objects under consideration (Sugitani et al. 2007).

16.3 Cellular Microfossils from Kaapvaal Craton, South Africa

The Kaapvaal Craton in the southern part of the African Shield is one of Earth’s major Archean cratons (Fig. 16.1a). Microfossils were reported from the Paleo- to Mesoarchean Barberton greenstone belt (Fig. 16.1b) and Late-Archean sedimentary units. The Barberton greenstone belt consists of a volcano-sedimentary succession called the Swaziland Supergroup. This supergroup comprises the 3.55–3.30 Ga Onverwacht, the 3.26–3.23 Ga Fig Tree, and the ~3.23 Ga Moodies Groups (Lowe 1999); microfossils and possible microfossils have been reported from all of these units (Table 16.2). Representative specimens from the Onverwacht Group and the Moodies Group are described.

16.3.1 *The Onverwacht Group*

Reports of fossil-like microstructures from the Onverwacht Group date back to the 1960s (e.g., Engel et al. 1968; Nagy and Nagy 1969); other publications followed in the 1970s (e.g., Muir and Hall 1974; Brooks et al. 1973). The structures included spheroids, filaments, and cup-shaped microstructures. However their biogenicity was not fully demonstrated, and thus they have generally been regarded as questionable or just as “possibility,” based on, for example, lack of clear colonial aggregations and wide size distribution (Schopf 1976; Schopf and Walter 1983). On the other hand, the biogenicity of carbonaceous filaments described from the Kromberg and Hooggenoeg formations (~3450 to 3334 Ma) has been assessed as being highly likely (Figs. 16.2a and 16.3a) (Walsh and Lowe 1985). Numerous curved and uniformly sized filaments were identified in carbonaceous chert of shallow-water origin. These filaments, especially tubular ones (from 1.4 to 2.2 μm in diameter and from 10 to 150 μm in length), were interpreted as being biogenic and having affinities with mat-forming bacteria or cyanobacteria (Walsh and Lowe 1985). Walsh (1992) also reported other morphological types, including spheroids, ellipsoids, and

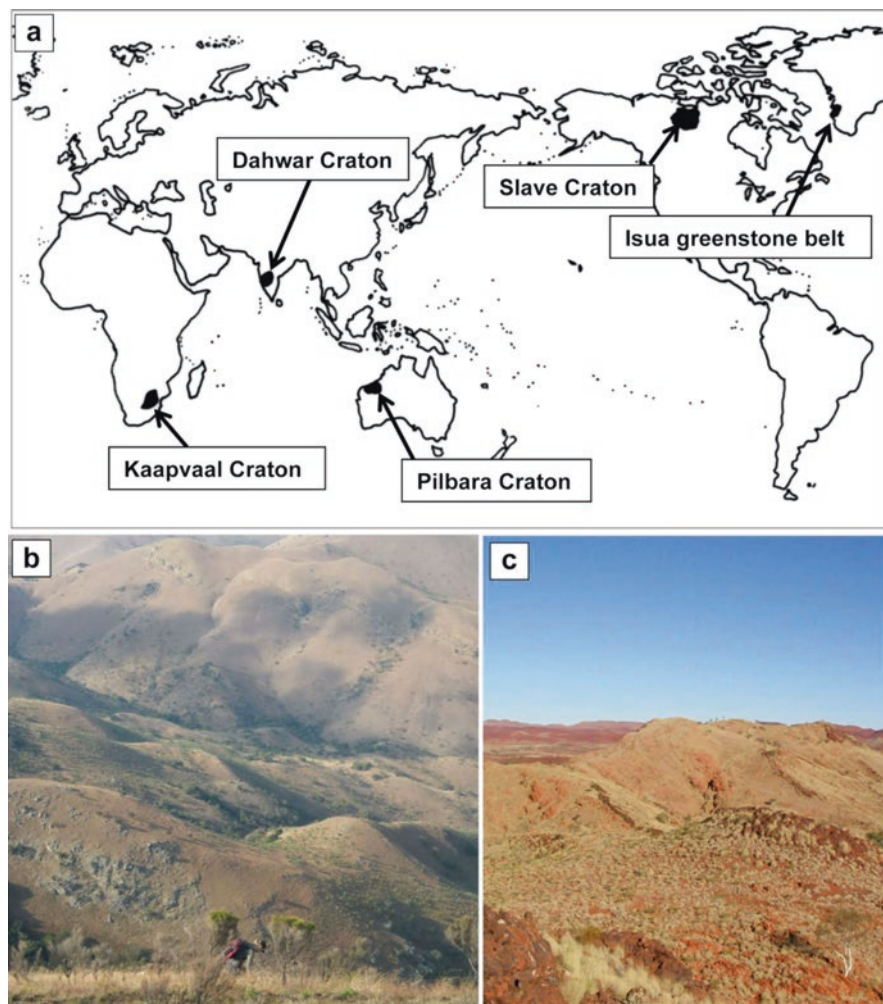


Fig. 16.1 (a) Representative cratons and greenstone belts containing Paleo- and Mesoarchean rocks. (b) Barberton greenstone belt in the Kaapvaal Craton, South Africa. (c) Panorama greenstone belt in the Pilbara Craton, Western Australia

spindles from the Kromberg Formation, and recognized several subtypes. Some of the microstructures are unexpectedly large for prokaryotic cells, which have cell sizes ranging mostly from 1 to 10 μm . Granular spheroids and ellipsoids are around 20 μm (up to 70 μm) in size (Fig. 16.3b), and spindles are around 30 μm (up to 140 μm) in length. The “spindles” were recently recollected and analyzed for their carbon isotopic compositions using SIMS (secondary ion mass spectrometry), and their morphology was reinterpreted as lenticular (Oehler et al. 2017) (Fig. 16.3c, d). It is noteworthy that morphologically equivalent microfossils were reported from the c. 3.4 Ga Strelley Pool Formation and the c. 3.0 Ga Farrel Quartzite in the Pilbara

Table 16.2 Paleo- and Mesoarchean (>3.0 Ga) microfossils (including putative ones) reported from the Kaapvaal Craton, South Africa

Formation (Group)	Maximum age (Ma)	Morphology	Depositional environment
Hooggenoeg Fm (Onverwacht Gp)	~3460	Narrow filaments, small spheroids, sausage-shaped structures (Walsh and Lowe 1985; Walsh 1992; Westall et al. 2001)	Shallow marine to subaerial setting (Walsh and Lowe 1985; Walsh 1992)
Kromberg Fm (Onverwacht Gp)	~3470	Diverse (tubular and solid) filaments, small spheroids and rods, large (~15 µm) spheroids, spindles, and lenses up to 150 µm long (Walsh and Lowe 1985; Walsh 1992; Westall et al. 2001; Tice and Lowe 2004; Oehler et al. 2017; Kremer and Kaźmierczak 2017)	Shallow marine, with low-temperature hydrothermal activity (Tice and Lowe 2004; Lowe and Worrell 1999; Hofmann and Harris 2008; Walsh 1992)
Swartkoppie Fm (Onverwacht Gp)	~3470	Small spheroids (Knoll and Barghoorn 1977)	Shallow marine (Knoll and Barghoorn 1977)
The Clutha Fm and not specified (Moodies Gp)	~3226	Large spheroids, molds of segmented thin filamentous structures (Javaux et al. 2010; Homann et al. 2016)	Shallow-water environments in tidal and deltaic settings, with microbial mats (Javaux et al. 2010; Heubeck et al. 2013)

Note: Prior to Knoll and Barghoorn (1977), putative microfossils, some from the Fig Tree Group, were described by Barghoorn and Schopf (1966), Schopf and Barghoorn (1967), Pflug (1967), Nagy and Nagy (1969), Engel et al. (1968), Brooks et al. 1973 and Muir and Hall (1974). They are not listed here because most of them were apparently misidentified as microfossils (e.g., Schopf and Walter 1983). Noteworthy is that some structures described in Pflug (1967) resemble lenticular microfossils from the Strelley Pool Formation and the Farrel Quartzite in the Pilbara Craton (see Table 16.3)

Craton, Western Australia (Sugitani et al. 2007, 2010). The significance of such occurrences will be discussed in more detail later. Kremer and Kaźmierczak (2017) also reported masses composed of small (3–12 µm) coccoid-like structures from the Kromberg Formation (Fig. 16.3e, f).

16.3.2 The Moodies Group

Javaux et al. (2010) reported spheroid microfossils from the younger Moodies Group (the Clutha Formation), which comprises siliciclastic rocks (shales and siltstone) deposited in tidal and deltaic settings. This finding is significant because siliciclastic rocks had been regarded as having low potential to preserve fragile microfossils. Additionally, the described microfossils are unusually large, up to about 300 µm in diameter, and are organic-walled (Fig. 16.4a, b). They were

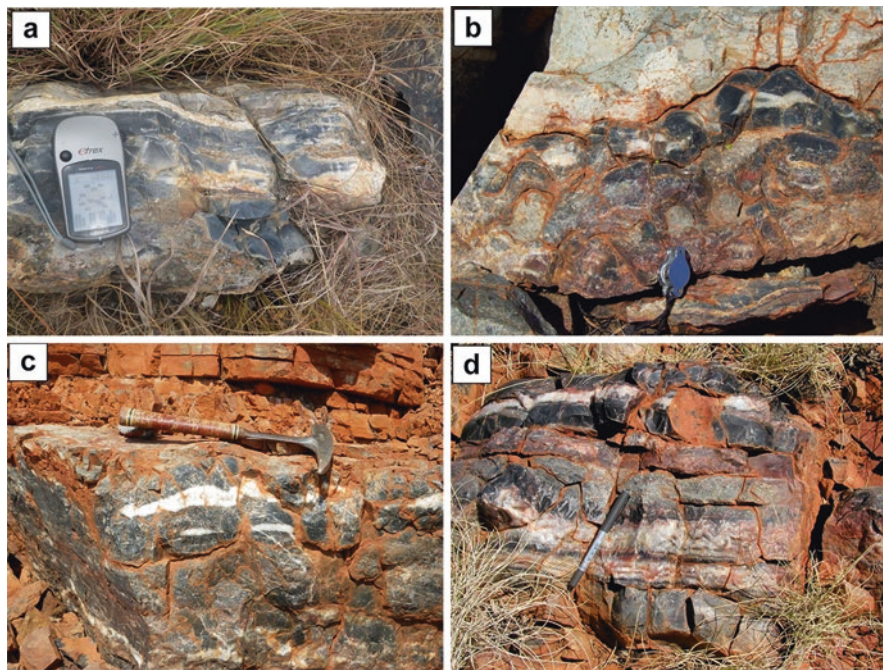


Fig. 16.2 Photographs of representative outcrops from which microfossils were discovered. (a) Black chert of the Kromberg Formation of the Onverwacht Group in the Kaapvaal Craton. (b) Black chert of the Strelley Pool Formation in the Goldsworthy greenstone belt of the Pilbara Craton. (c) Black chert of the Strelley Pool Formation in the Panorama greenstone belt of the Pilbara Craton. (d) Black chert of the Farel Quartzite in the Goldsworthy greenstone belt of the Pilbara Craton

identified in petrographic thin section, as well as in macerates from acid digestion: gentle decomposition at room temperature using hydrochloric acid (HCl) and hydrogen fluoride (HF). Javaux et al. (2010) discussed their biological affinities, comparing them to large extant prokaryotes.

Abundant filamentous microfossils preserved as molds were also reported from the Moodies Group (Homann et al. 2016). These authors examined lens-shaped chert with kerogenous laminae comprising downward-growing microstromatolitic columns in siliciclastic rocks and identified abundant filamentous structures using scanning electron microscope (SEM). The cylindrical and hollow microfilaments from 0.3 to 0.5 μm in diameter are bent in various directions and are regularly segmented (Fig. 16.4c, d). Such features, together with the carbon isotopic values of associated kerogenous laminae ($\delta^{13}\text{C}_{\text{PDB}} = -26.5$ per mil on average), are consistent with a biogenic origin. The lens-shaped chert was interpreted as representing an early silicification of cavities formed beneath microbial mats. The traces of life described may be evidence of the presence of cavity-dwelling microbes protecting themselves from the supposed intense ultraviolet radiation at that time (Homann et al. 2016).

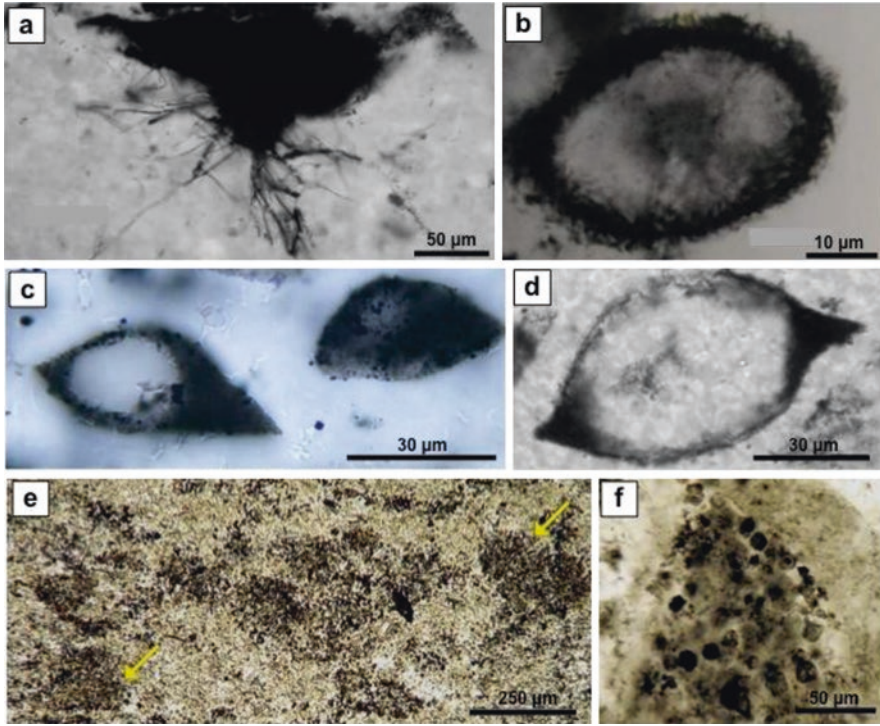


Fig. 16.3 Photomicrographs of microfossils from the Onverwacht Group, South Africa. (a) Filamentous microfossils from the Kromberg Formation (Walsh and Lowe 1985) (Photo courtesy by Maud Walsh). (b) Granular-walled, large spheroid from the Kromberg Formation (Walsh 1992) (Photo courtesy by Maud Walsh). (c, d) Lenticular microfossils from the Kromberg Formation discovered independently by M.M. Walsh and K. Sugitani, respectively (Oehler et al. 2017, with permission from Elsevier). (e) Subglobular to spindle clusters (arrows) containing abundant cell-like bodies (Kremer and Kaźmierczak 2017, with permission from Elsevier). (f) Magnified image of cluster of cell-like bodies, interpreted as the remains of variously degraded colonies of cyanobacterial-like microbes. (Kremer and Kaźmierczak 2017, with permission from Elsevier)

16.4 Cellular Microfossils from Pilbara Craton, Western Australia

The Pilbara Craton in northwestern Australia is composed of the East Pilbara, Regal, Karratha, Sholl, and Kurrana Terranes, collectively overlain by the De Grey Supergroup (Fig. 16.1). The oldest rocks of the craton are preserved in the East Pilbara Terrane, where greenstone belts assigned to the Pilbara Supergroup consist of a series of four major groups and one separate formation (Hickman 2012; Van Kranendonk et al. 2002, 2006, Hickman and Van Kranendonk 2008). Most microfossils and possible microfossils have been reported from this supergroup, including the c. 3.52–3.42 Ga Warrawoona Group, the c. 3.4 Ga Strelley Pool Formation, the c. 3.24 Ga Sulphur Springs Group, and the c. 3.0 Ga Gorge Creek Group of the

Table 16.3 Paleo- and Mesoarchean (>3.0 Ga) microfossils (including putative ones) reported from the Pilbara Craton, Western Australia

Formation (Group)	Maximum age (Ma)	Microfossils	Depositional environment
Dresser Fm (Warrawoona Gp)	3490	Small spheroids, threads, septate or, nonseptate filaments (Dunlop et al. 1978; Ueno et al. 2001a, b; Glikson et al. 2008)	Various depositional environments have been proposed for the Dresser Fm (shallow to subaerial and temporary evaporitic shallow marine, oceanic caldera, deep-sea hydrothermal vent, terrestrial hydrothermal system, coastal sabhka) (Isozaki et al. 1997; Buick and Dunlop 1990; Van Kranendonk 2006, 2007; Van Kranendonk et al. 2008; Noffke et al. 2013; Djokic et al. 2017)
Mount Ada Basalt (chert) (Warrawoona Gp)	3470	Septate or nonseptate filaments, small spheroids (Awramik et al. 1983)	Shallow marine? Not confirmed
Apex Basalt (chert) (Warrawoona Gp)	3460	Septate or nonseptate filaments, spheroids (Schopf and Packer 1987; Schopf 1993; Schopf et al. 2018)	Shallow marine (Schopf 1993)? High temperature hydrothermal vent system (Brasier et al. 2002)?
Panorama Fm (Warrawoona Gp)	3446	Small spheroids, small rods, filaments, films (Westall et al. 2006, 2011)	Shallow marine, with low-temperature hydrothermal activities (DiMarco and Lowe 1989)
Strelley Pool Fm	3426	Small to large spheroids, lenses, films, diverse filaments (Sugitani et al. 2010, 2013, 2015a, b; Schopf et al. 2017; Oehler et al. 2017)	Shallow marine (intertidal to supratidal) setting, with hydrothermal inputs (Hickman 2008; Sugitani et al. 2010, 2013; Allwood et al. 2007, 2010), locally possibly terrestrial hydrothermal coastal field (Sugitani et al. 2015b)
Kangaroo Caves Fm	3240	Filaments, small spheroids (Rasmussen 2000; Duck et al. 2007)	Deep marine (>1500 m), high temperature hydrothermal setting (Vearncombe et al. 1995)
Dixon Island Fm	3200	Small spheroids, septate, or nonseptate filaments (Kiyokawa et al. 2006)	Relatively deep marine, hydrothermal setting (Kiyokawa et al. 2006, 2014)
Farrel Quartzite (Gorge Creek Gp)	3050	Small to large spheroids, lenses, diverse films, filaments (Sugitani et al. 2007, 2009a, b; Grey and Sugitani 2009)	Shallow marine, possibly a closed to semi-closed basin, with input of continental runoff and/or low-temperature hydrothermal fluids (Sugitani et al. 2003; Sugahara et al. 2010)

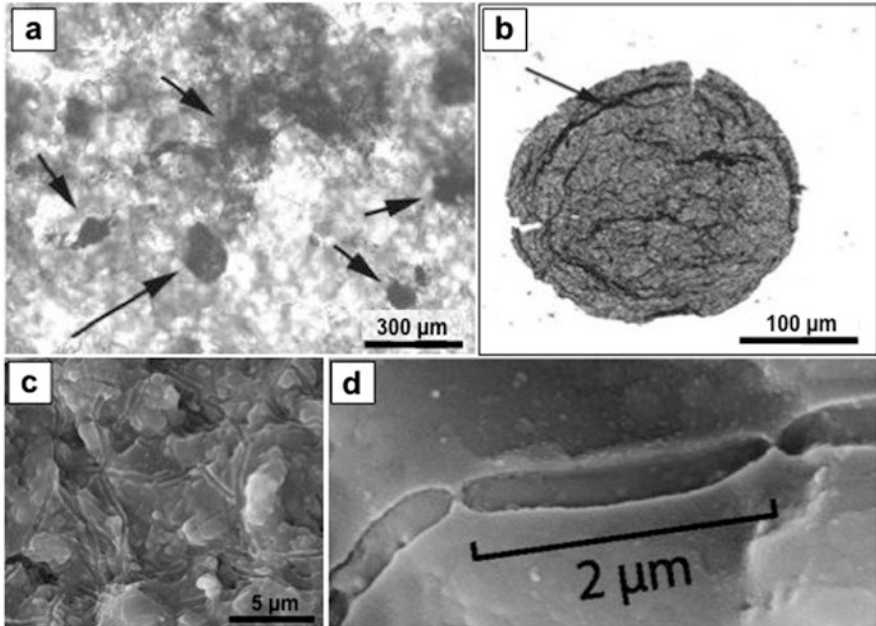


Fig. 16.4 Microfossils from the Moodies Group, South Africa. **(a, b)** Carbonaceous spheroid microfossils in siliciclastic rock (shale/siltstone) (Javaux et al. 2010, with permission from Springer). **(a)** Abundant compressed carbonaceous spheroids (arrows) observed in petrographic thin section. **(b)** Specimen extracted by acid maceration. The arrow shows a concentric fold. **(c, d)** Secondary-electron photomicrographs of filamentous microfossils in silicified cavities below microbial mat within sandstone (Homann et al. 2016) (Photo courtesy by Martin Homman). **(c)** Meshwork of filamentous molds embedded in chert. **(d)** An enlargement image of **(c)**, showing regularly spaced, rod-shaped segments

overlying De Grey Supergroup. In addition, the c. 3.2 Ga Dixon Island Formation in the Western Pilbara contains possible microfossils (Table 16.3). In the following discussion, representative specimens of especial significance reported from the Dresser Formation and the Mount Ada Basalt of the Warrawoona Group, the Strelley Pool Formation and the Farrel Quartzite of the Gorge Creek Group are reviewed.

16.4.1 Dresser Formation of the Warrawoona Group

The Dresser Formation is predominantly composed of blue, black, and white layered chert, minor carbonate rocks, and barite, with komatiitic basalt (Van Kranendonk et al. 2008). Buick and Dunlop (1990) suggested that it was deposited in a closed to semi-closed shallow coastal basin. On the other hand, Van Kranendonk (2006) and Van Kranendonk et al. (2008) claimed that it represented an ancient active volcanic caldera, with at least temporal subaerial exposure during which geysers were

deposited (Djokic et al. 2017). Hydrothermal origin has been previously proposed by Isozaki et al. (1997), who however claimed deep-sea mid-ocean ridge setting. Terrestrial environment (sabhka) was also suggested by Noffke et al. (2013). Dunlop et al. (1978), Ueno et al. (2001a, b), and Glikson et al. (2008) reported microfossils and possible microfossils.

Dunlop et al. (1978) extracted tiny, hollow carbonaceous spheroids from carbonaceous cherts by acid maceration. Spheroids 1.2–12 μm in diameter were identified. They included solitary spheroids, spheroids with splits, paired spheroids, and rare chains of spheroids (Fig. 16.5a). However, these structures are now considered to be viscous bitumen droplets or mineralic non-biological spheroids (Buick 1990; Awramik et al. 1983; Schopf et al. 1983; Wacey 2009).

Ueno et al. (2001a, b) reported filamentous microstructures from both hydrothermal silica dykes and bedded cherts in the Dresser Formation, describing solitary spiral filaments, radiating clusters of threads, and tubular filaments $\sim 10 \mu\text{m}$ thick (Fig. 16.5b). Carbon isotopic values of these structures were individually analyzed using SIMS (Ueno et al. 2001a). The obtained significantly light ($\delta^{13}\text{C} < -30$ per mil) values were interpreted as possible products of the Calvin cycle or reductive acetyl-CoA pathway. Glikson et al. (2008) performed acid maceration on bedded black chert samples from the Dresser Formation. Several types of carbonaceous objects including tiny spheroid bodies (less than 3 μm to submi-

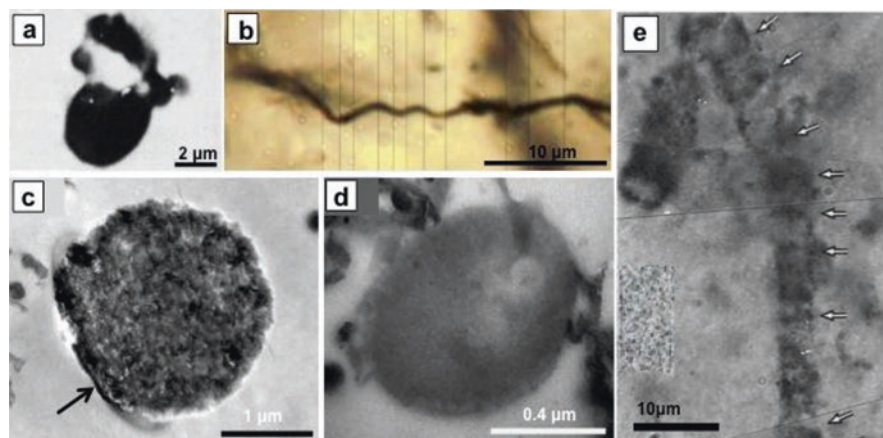


Fig. 16.5 Microfossils and microfossil-like structures from the Dresser Formation and the Mount Ada Basalt of the Warrawoona Group in the Pilbara Craton. (a) Probable pseudomicrofossil, carbonaceous spheroid extracted by acid maceration showing ruptured split (Dunlop et al. 1978, with permission from Springer). (b) Sinuous carbonaceous filament in black chert (Ueno et al. 2001a) (Photo courtesy by Y.Ueno). (c, d) TEM images of possible cells (Glikson et al. 2008, with permission from Elsevier). (c) Specimen with a partial preserved cell-like wall segment (arrow). (d) Elliptical cell-like structure entombed within fluid inclusions. (e) Acute folded relative thick septate filament from chert of the Mount Ada Basalt (Awramik et al. 1983, with permission from Elsevier). The arrows show septation

cron in diameter) (Fig. 16.5c, d) were obtained, though it is equivocal whether such simple and tiny spheroids were certainly originated from cells. Also carbonaceous objects associated with the spheroids were interpreted as separated cell walls, equivalents to thermally degraded cells of the living hyperthermophile *Methanocaldococcus jannaschii*. Evidence for microbial methanogenesis in this formation was also obtained from analyses of fluid inclusions (Ueno et al. 2006).

16.4.2 *Mount Ada Basalt of the Warrawoona Group*

Awramik et al. (1983) described carbonaceous spheroids and filaments in bedded chert collected from the Mount Ada Basalt in the North Pole area. A prokaryotic affinity for some tubular and septate filamentous structures was suggested. One of the described filaments is relatively thick (~5 μm), acutely folded, and regularly septate (Fig. 16.5e). Such distinct features point to its biogenicity. Buick (1984), on the other hand, questioned whether the host chert samples were of primary sedimentary origin, whether they were from the Mount Ada Basalt or a much younger unit (Fortescue Group), and whether the microstructures described are microfossils in Archean age, leading to a published debate (Buick 1988; Awramik et al. 1988). Uncertainty about the sample locality was later resolved, and the host cherts were indeed in the Mount Ada Basalt (Grey et al. 2010), although recollection of similar materials has not yet been successful.

16.4.3 *Strelley Pool Formation*

The Strelley Pool Formation (Fig. 16.1c), typically 8–11 m but locally up to 100 m thick, is composed of siliciclastic sedimentary rocks, bedded and stromatolitic dolomite, chert, and volcanoclastic rocks. The depositional environment ranges widely from a fluvial to deeper water setting and includes a rocky coastal shoreline and beach environment, a shallow-water marine carbonate platform, a possible sebkha, a tidal flat, an alluvial fan, and a possible terrestrial hydrothermal system (Allwood et al. 2006, 2007; Hickman 2008; Van Kranendonk, 2011; Sugitani et al. 2015b).

Schopf and Packer (1987) and Schopf (2006) reported spheroid structures in carbonaceous chert from the Strelley Pool Formation. The structures are composed of sheathed multiple spheroids, similar to chroococcacean cyanobacteria and are up to over 50 μm across. Their biogenicity is likely considering such morphological elaboration (Schopf 2006).

Compelling evidence for microfossils in this formation was later presented by Sugitani et al. (2010, 2013, 2015a), who described morphologically diverse carbonaceous microstructures in chert from three widely separated (~50 km) localities in the Panorama, Warralong, and Goldsworthy greenstone belts (Fig. 16.2b, c). The microfossil assemblages are composed of spheroidal, lenticular, and filamentous

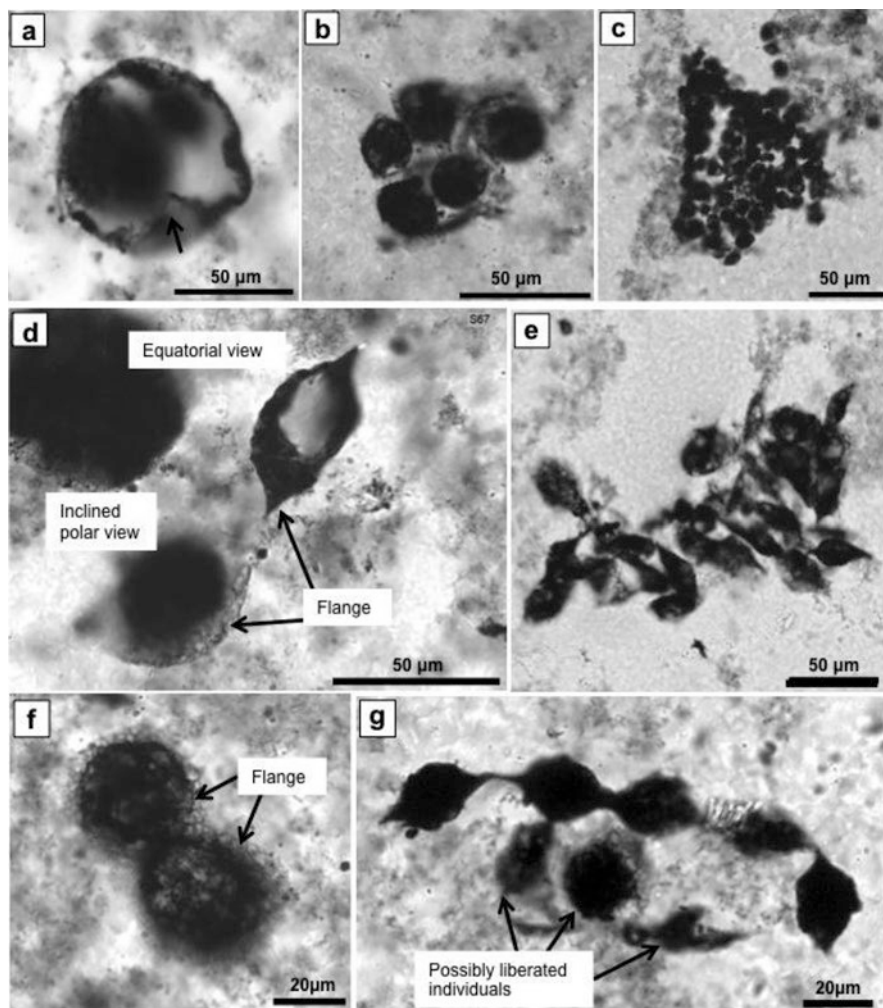


Fig. 16.6 Carbonaceous microfossils from the Strelley Pool Formation in the Pilbara Craton. (a) Large ($>50\ \mu\text{m}$ in diameter) hollow spheroid with a partial broken wall (arrow). (b) Colony composed of relatively tightly packed spheroids $\approx 20\ \mu\text{m}$ in diameter. (c) Colony composed of abundant small spheroids $5\text{--}15\ \mu\text{m}$ in diameter. (d) Two specimens of lenticular microfossil preserved in different orientations in petrographic thin section. (e) Colony composed of abundant, randomly directed, lenticular microfossils. (f) Polar view of paired lenticular microfossils. (g) Chain composed of lenticular microfossils, with individuals possibly liberated from the chain

microfossils. In addition, film-like carbonaceous objects were identified and interpreted as fragmented biofilm.

Spheroidal microfossils range from $5\ \mu\text{m}$ up to $60\ \mu\text{m}$ in diameter but are mostly less than $15\ \mu\text{m}$ (Fig. 16.6a–c). They are solitarily or comprise colony-like clusters.

Lenticular microfossils range mostly from 20 to 60 μm along the major dimension in polar view and rarely up to 100 μm . The structures are characterized by a thin, sheet-like appendage or flange surrounding the central body (Fig. 16.6d) (Sugitani et al. 2007). Microfossils of this morphological type show minor morphological variations, including ellipticity in polar view, transparency of the central body, and variations in flange width and texture (Sugitani et al. 2010, 2013, 2015a, 2018). They can be solitary or comprise colony-like clusters (Fig. 16.6e). Paired or chained lenticular microfossils are also present (Fig. 16.6f, g).

Filamentous microfossils include thin threads and hollow tubes. Hollow tubes up to 200 μm in length and up to 20 μm in width show the most compelling evidence for biogenicity (Fig. 16.7a, b).

The biogenicity of the Strelley Pool spheroidal and lenticular microstructures is well established. Their carbonaceous composition, size distribution, taphonomic features, and colonial-like associations are all consistent with a biogenic origin (Sugitani et al. 2010, 2013). This is supported by their light (< -30 per mil), texture-specific carbon isotopic values (Lepot et al. 2013) and by the alveolar or hollow internal texture of the central body as revealed by SEM and transmission electron microscopy (TEM) (Sugitani et al. 2015a). The Strelley Pool lenticular microfossils and their composite structures (pairs and chains) can be extracted by acid maceration, which confirms that they are organic-walled microfossils.

Wacey et al. (2011) described carbonaceous spheroids, ellipsoids, and tubular filaments associated with pyrite crystals from the basal sandstone of the Strelley Pool Formation in the East Strelley greenstone belt (Fig. 16.7c, d). The structures have carbonaceous matter in the walls, with $\delta^{13}\text{C}$ values of -33 to -46 ‰, associated with nitrogen. They also show taphonomic degradation and organization into chains and clusters. The authors suggested an affinity with sulfur-metabolizing bacteria (Wacey et al. 2011). Recently, other types of microfossils with possible affinities to sulfur-metabolizing bacteria were reported from the Panorama greenstone belt (Schopf et al. 2017). These are from the same locality as for the lenticular microfossils first described by Sugitani et al. (2010). Schopf et al. (2017) described carbonaceous cobweb-like structures and associated thin tubular microfossils (a few microns wide) in carbonaceous chert (Fig. 16.7e, f). By analogy with younger equivalents (Schopf et al. 2015) and from sulfur isotopic analyses, the authors suggested that they represent fossilized sulfur bacterium.

16.4.4 *Farrel Quartzite of the Gorge Creek Group*

The Farrel Quartzite (Van Kranendonk et al. 2006) in the Goldsworthy greenstone belt is dominated by very coarse-grained sandstone, including quartzite and minor conglomerate, with minor amounts of mafic to ultramafic volcanoclastic beds, evaporite beds, and layers of black chert. The total thickness varies along strike from several meters up to 80 m (Sugitani et al. 2003, 2007). Microfossils were discovered

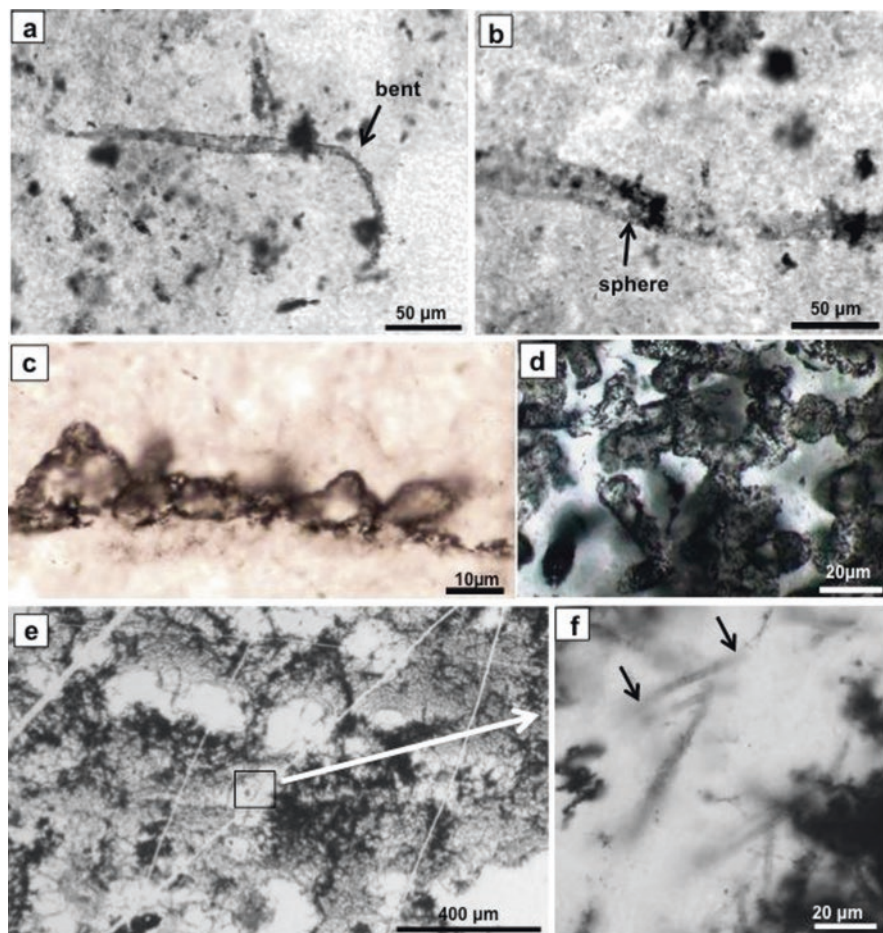


Fig. 16.7 Carbonaceous microfossils from the Strelley Pool Formation in the Pilbara Craton. (a) Bent tubular carbonaceous filament. (b) Highly degraded possible tubular filament containing small sphere inside (arrow). (c) Spheroids and (d) tubular filaments in silica-filled pore spaces in the basal sandstone (Wacey et al. 2011) (Photo courtesy by David Wacey). (e) Cobweb-like structure in carbonaceous black chert partially containing filamentous microfossils (Schopf et al. 2017). (f) Narrow tubular filamentous microfossils. Their sinuous morphology is indicated by blurred terminals (arrows) (Schopf et al. 2017)

in black chert in the uppermost unit of the Farrel Quartzite (Sugitani et al. 2007) (Fig. 16.2d). The microfossiliferous chert layer can be traced for c. 7 km along strike. Sugitani et al. (2007) established an aqueous deposition for the chert because of the presence of spherulitic structures indicative of silica gel precipitation and cross-lamination identified from the distribution of carbonaceous objects indicative of transportation by fluid flows. Sugahara et al. (2010) also showed that the rare-earth element yttrium data of the cherts could best be interpreted as mixing of

seawater and low-temperature hydrothermal fluids and/or terrestrial runoff. Retallack et al. (2016), on the other hand, reinterpreted the fossil-bearing black chert as paleosol, which was refuted in detail by Sugitani et al. (2017).

The Farrel Quartzite microfossil assemblage includes morphologically diverse carbonaceous structures such as spheroids, lenses, and films that are basically similar to those in the Strelley Pool Formation assemblage. Spheroids range from <5 to 80 μm in diameter. Smaller (<15 μm) spheroids are very abundant, and their populations can be ~500 in a thin section (2.5 \times 3.4 cm, with 30 μm in thickness). Such small spheroids often comprise colony-like clusters (Fig. 16.8a–c). Large spheroids display various wall textures such as dimpled, folded, and fluffed (Fig. 16.8d–f). Some of the large, hollow spheroids contain multiple small spheroids or a single spheroid inside (Fig. 16.8g, h). Colonies composed only of larger (>20 μm) spheroids are also present but not common, whereas colonies composed of spheroids of different sizes, including large ones, are relatively common (Fig. 16.8i).

Like the lenticular microfossils in the Strelley Pool Formation, the Farrel Quartzite lenses are composed of central body with a surrounding flange (see Fig. 16.6d for reference). The whole structure ranges from 20 to >100 μm along the maximum dimension. Lenses occur solitarily, as pairs, or comprising colonies. Colonies containing lenses are composed either exclusively of lenses like the Strelley Pool lenses (Fig. 16.6e, g) or of mixture of lenses and spheroids (Fig. 16.9a, b). Sugitani et al. (2007, 2009a) identified minor morphological variations in lenticular microfossils. The central body is either symmetric or asymmetric in equatorial view (Sugitani et al. 2007; Fig. 15). Asymmetry is identified along the long axis, the short axis, or both. In some specimens, the width of the flange is also asymmetrical in polar view. Although rare, even more highly elaborate specimens have been recorded, such as a lens with an inner sphere, a lens apparently expelling a sphere (or lens), and a lens containing plural inner bodies (Fig. 16.9c–e). Lenticular microfossils also display variations in flange texture, ranging from translucent to hyaline, reticulate, and striated (Sugitani et al. 2009a) (Fig. 16.9f–h).

Biogenicity of the Farrel Quartzite carbonaceous structures was compellingly demonstrated by Sugitani et al. (2007, 2009a), Oehler et al. (2009, 2010), Grey and Sugitani (2009), Schopf et al. (2010), and House et al. (2013). Lines of evidence include (1) geologic context suitable for life, (2) carbonaceous composition and carbon isotopic values of individual specimens, (3) association of nitrogen and sulfur with the kerogenous structure, (4) the presence of complex internal reticulate structures (for lenticular structures), (5) flexibility and breakability as an original physical property giving rise to taphonomic features, (6) abundance of the structures and their relatively narrow size distributions, (7) complex occurrences such as the coexistence of different morphological types in colonies, and (8) elaborate morphologies suggestive of reproduction. Grey and Sugitani (2009) also proved that lenses, spheroids, and films are acid-resistant and could be extracted by conventional acid maceration of the host rocks using hydrochloric and hydrofluoric acids (Grey 1999), which produces an almost exclusively organic residue.

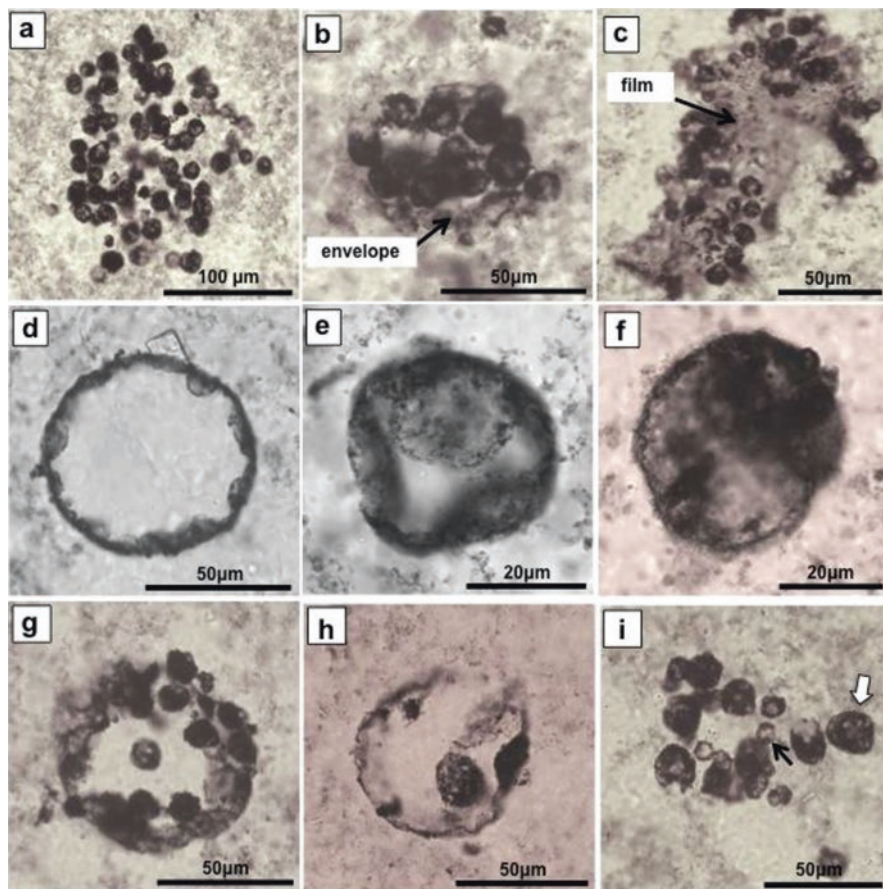


Fig. 16.8 Carbonaceous spheroid microfossils from the Farrel Quartzite. (a) Colony composed of loosely packed small (<15 μm) spheroids. (b) Colony of relatively tightly packed small spheroids, with possible remains of an envelope (arrow). (c) Small spheroids attached to a film-like structure (possible microbial mat remnants). (d) Large (>50 μm), hollow spheroid with dimpled wall (the dimples are possible fold structures). (e) Large (>30 μm) spheroid with highly folded wall. (f) Large (>40 μm) spheroid with a fluffy wall. (g) Large spheroid with dispersed, small (~10 μm), internal spheroids. (h) Ruptured large spheroid with an inner spheroid. (i) Colony composed of spheroids of highly different sizes. Specimens of contrasting sizes are arrowed

16.5 Discussion

16.5.1 Implications for Early Evolution of Photosynthesis

Depositional environments for Paleo- and Mesoarchean microfossil-bearing sedimentary rocks from the Kaapvaal and the Pilbara cratons are summarized in Tables 16.2 and 16.3. They show that microbes were already flourishing in a variety of

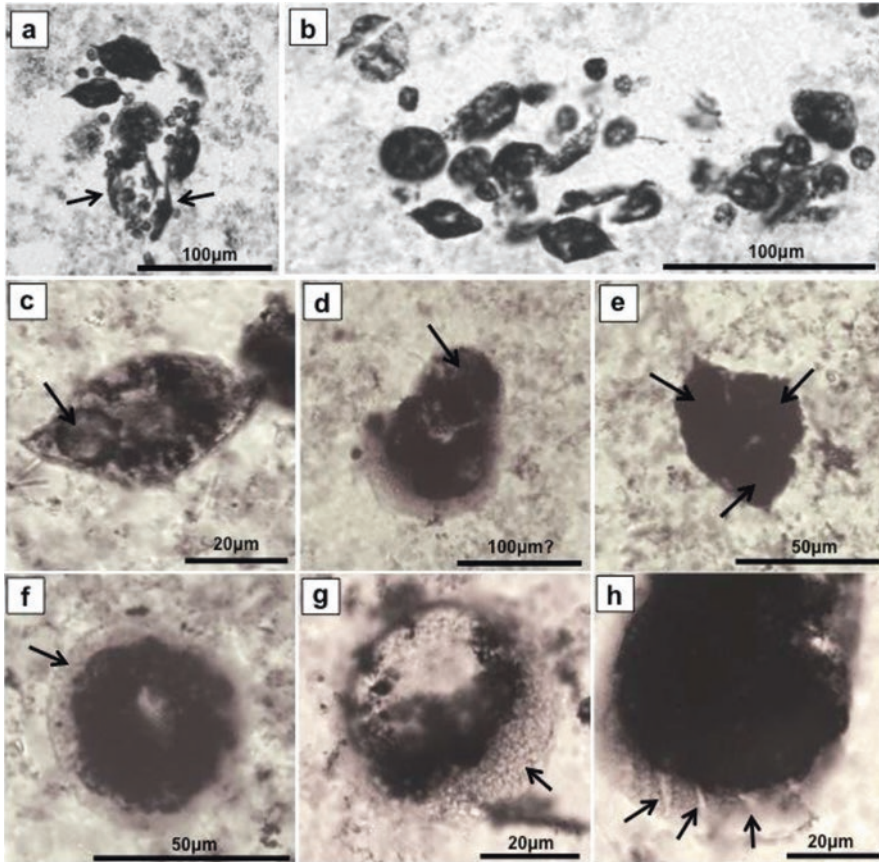


Fig. 16.9 Carbonaceous lenticular microfossils from the Farrel Quartzite. (a) Colony composed of lenticular and small spheroid microfossils. (b) Colony composed of lenticular and spheroid microfossils of various sizes. (c) Lenticular microfossil containing a small hollow sphere inside. (d) Polar view of lenticular microfossil with an equidimensional flange (polar view) apparently expelling a slightly smaller object. (e) Lenticular microfossil that appears to contain inner three bodies (arrows). (f) Polar view of lenticular microfossil with translucent and hyaline flange. (g) Lenticular microfossil with reticulated flange. (h) Polar view of lenticular microfossil with broad partial striations (the arrows)

environments well before 3.0 Ga (Van Kranendonk 2011), mostly in shallow to subaerial environments. This is consistent with records of non-microfossiliferous biosignatures, such as stromatolites and microbially induced sedimentary structures (e.g., Hofmann et al. 1999; Allwood et al. 2006; Djokic et al. 2017; Noffke et al. 2013). This may provide some hints about the metabolism of ancient microorganisms. An ozone shield was unlikely to have been present, considering the very low concentrations of free oxygen during the Archean (e.g., Holland 2006, Chap. 17).

Shallow-water to subaerial habitats were probably radiated by harmful ultraviolet (UV) rays, although the lack of an ozone layer might have been compensated for by a UV-shielding organic haze (Wolf and Toon 2010), dispersed silica particles in water columns (Siever 1992; Stefurak et al. 2014), and/or the encrustation of cells with silica (Phoenix et al. 2001). In either case, one of the benefits from living in such environments would have been the utilization of light energy. This suggests that photosynthesis and even oxygenic photosynthesis might have already emerged in the Paleoarchean, as argued by or suggested from previous studies (Buick 2008; Nisbet and Sleep 2001; Tice and Lowe 2004, 2006a, b; Hoashi et al. 2009; Mukhopadhyay et al. 2014; Rosing and Frei 2004; Crowe et al. 2013; Planavsky et al. 2014; Lyons et al. 2014; Schirmermeister et al. 2016). The discovery of an axial zone in the large conical stromatolites in the Strelley Pool Formation at the Trendall Geoheritage Reserve (Hickman et al. 2011) may also support the idea that oxygenic photosynthesis was occurring locally, because modern motile cyanobacteria in hot springs have been shown to move toward light and in so doing to produce a similar axial-zone structure (Walter et al. 1976). Recently, Kremer and Kaźmierczak (2017) suggested that small spheroids in the 3.4 Ga Kromberg Formation represent fossilized coccoid cyanobacteria such as *Microcystis*. While such an early evolution of oxygenic photosynthesis is not always widely accepted and needs further studies, the emergence of anoxygenic photosynthesis is likely back to Paleoarchean (e.g., Tice and Lowe 2004).

16.5.2 Taxonomic and Phylogenic Implications of Paleo- and Mesoarchean Filamentous Microfossils

Filamentous microfossils with some morphological variations have been reported from the Pilbara and the Kaapvaal cratons, and to my knowledge, the first reliable report can be back to the 1980s (Walsh and Lowe 1985). Although there have been some debates on biogenicity of earlier reported Archean filamentous structures (e.g., Buick 1984), recent new discoveries and revisits to the previously described specimens have confirmed the presence of this morphotype of microbes in the Paleo- and Mesoarchean (Homann et al. 2016; Grey et al. 2010; Schopf et al. 2017, 2018; Sugitani et al. 2013). Furthermore taxonomic diversity of filamentous microfossils can be inferred from observed morphological variations in, e.g., length, diameter, and presence or absence of septation. Although these filamentous microfossils are likely prokaryotic, further assignments to specific extant taxa require careful examinations. Finally, the recent discoveries of tubular filaments locally well-preserved in carbonaceous cherts with biomat-like fabrics (Fig. 16.7f) (Schopf et al. 2017) and of cherty stromatolite preserving carbonaceous laminae (Sugitani et al. 2015b) suggest possibility of future discovery of cellularly preserved stromatolite builders.

16.5.3 Taxonomic and Phylogenic Implications of Paleo- and Mesoarchean “Large” Microfossils

This section focuses on large (> ca. 20 μm) microfossils with a spheroidal or lenticular shape from both the Pilbara and Kaapvaal cratons (e.g., Sugitani et al. 2007, 2010; Javaux et al. 2010; Oehler et al. 2017). The size of a fossilized cell could place some constraints on their biological affinity, although this is not diagnostic. In general, prokaryotic coccoid or rod-shaped cells are smaller than 10 μm . Many of the newly described spheroidal and lenticular microfossils range from 20 to 300 μm in maximum diameter and are therefore extremely and uncharacteristically large for prokaryotic cells. Additionally, they are organic-walled and can be extracted by acid maceration. Namely, they can have acid-resistant recalcitrant wall or envelope. This ability is known widely for eukaryotes, although some prokaryotes have acid-resistant cell walls. In addition, it should be remembered that some extant prokaryotes can be as large as single-celled eukaryotes. How can we explain such extremely large sizes? What are their biological affinities? These questions are difficult to answer at present but could be resolved as morphological and chemical data improves and as the differences and similarities between extant prokaryotes and eukaryotes become clearer.

16.5.3.1 Large Spheroids

Large (> ca. 20 μm) Paleo- and Mesoarchean spheroids have been reported from the 3.4 Ga. Strelley Pool Formation, the 3.4 Ga Kromberg Formation, the 3.2 Ga Moodies Group, and the 3.0 Ga Farrel Quartzite, as described earlier. Specimens from the Farrel Quartzite and the Moodies Group have been proved to be organic-walled and acid-resistant (Grey and Sugitani 2009; Javaux et al. 2010). Specimens from the Strelley Pool and Kromberg Formations and the Farrel Quartzite are up to 80 μm in diameter, whereas those from the Moodies Group can even reach 300 μm in diameter. The size distribution differences may be of taxonomic significance. The relatively smaller three populations occur in carbonaceous chert deposited in a shallow-water environment with hydrothermal influence, whereas the other occurs in siliciclastic sediments deposited in tidal to deltaic environments (Tables 16.2 and 16.3), although further studies are needed to confirm this possibility.

When considering the biological affinity of spheroids larger than 20 μm in diameter, it must be kept in mind that some prokaryotic coccoid microbes can reach similar large sizes. For example, some extant spheroidal cyanobacteria such as *Dermocarpella* and *Staniera* (Waterbury and Stanier 1978; Angert 2005) can be 20 μm or more in maximum diameter. These large cyanobacteria are characterized by vegetative cell growth and reproduction by multiple fissions (e.g., Angert 2005) (Fig. 16.10a). Spheroidal sulfur bacteria, *Thiomargarita namibiensis*, can reach 750 μm in diameter and can form chains (Fig. 16.10b), although it should be noted

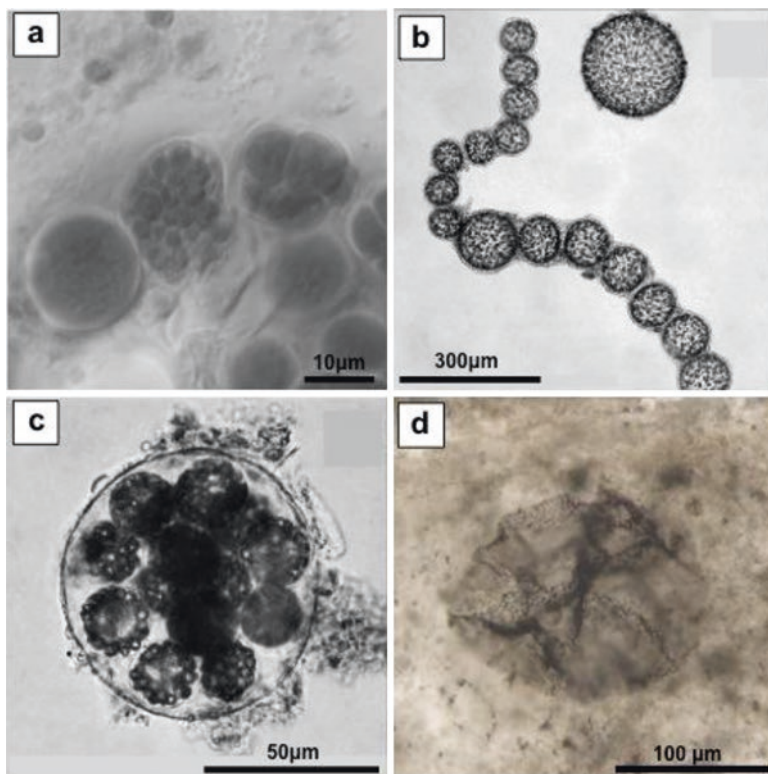


Fig. 16.10 (a) Cyanobacteria containing numerous baeocytes (the center), likely related to genus *Chroococcopsis* (photo courtesy by Sergei Shalygin). (b) Large chain-forming sulfur bacteria *Thiomargarita namibiensis* (Salman et al. 2011) (with permission from Elsevier). (c) Colony-forming sulfur bacteria (cf. *Thiomargarita nelsonii*) with a surrounding rigid envelope (Salman et al. 2013) (with permission from Elsevier). (d) Carbonaceous spheroidal microfossil (in petrographic thin section) from the 2.52 Ga Gamohaan Formation of South Africa (Czaja et al. 2016). (Photo courtesy by A.D. Czaja). Note this spheroid can be extracted by acid maceration

that the cytoplasm exists only peripherally and their cells are volumetrically dominated by a vacuole (Schulz et al. 1999; Salman et al. 2011).

A few specimens of large spheroids from the Farrel Quartzite contain small internal spheroids, and colonies composed of spheroids of different sizes are relatively common (Fig. 16.8g, i). Based on these observations, Sugitani et al. (2009b) suggested that the large spheroids might represent spheroidal cyanobacteria reproducing by multiple fissions (Fig. 16.10a). However, the possibility that they were sulfur bacteria cannot be excluded, especially considering that a newly identified taxon of sulfur bacteria (Ca. *Thiomargarita nelsonii*) includes specimens that form a colony of cells surrounded by a rigid envelope (Salman et al. 2011, 2013) (Fig. 16.10c), which is similar to the Farrel Quartzite specimen illustrated in Fig. 16.8g.

Javaux et al. (2010) discussed the biological affinity of the Moodies Group large, organic-walled large spheroids in the context of their ecological niches. Some candidates for large extant prokaryotes, such as the fish gut, parasitic *Epulopiscium* (Angert 2005) and the sulfur bacteria, were discounted, because their ecological niches are totally different from that of the Moodies Group large microfossils. Moreover, they have no ability to produce recalcitrant biopolymers. On the other hand, they suggested a possible affinity to cyanobacteria because some cyanobacteria produce large cysts and envelopes as described previously and because the Moodies Group habitat was in the non-sulfidic photic zone.

Additionally, microfossils reported from a younger succession, the 2.52 Ga Gamohaam Formation of the Kaapvaal Craton, a deepwater facies of a carbonate platform (Czaja et al. 2016) (Fig. 16.10d), require consideration. Czaja et al. (2016) described organic-walled microfossils up to 265 μm in diameter that have reticulated walls and display compression features consistent with sediment compaction. The authors suggested a possible affinity to sulfur-oxidizing bacteria, such as *Thiomargarita*, based on their environment, morphological features (size and shape), and sulfur isotopic signatures.

Finally, it should be noted that large spheroids have significant morphological diversity, which might be expressions of biotic diversity. As described earlier, variations in wall texture (Fig. 16.8d–f) cannot be explained only by taphonomy. Additionally, large spheroids do not always represent enlarged mother cells or simply large vegetative cells. For example, some specimens that contain a single, internal sphere are more probably explained as endospore enclosed in a recalcitrant envelope (Fig. 16.8h) (Sugitani et al. 2009b).

16.5.3.2 Lenses

Lenses (lenticular microfossils) were first described as spindles (spindle-like structures) in Sugitani et al. (2007), following the term given to similar structures discovered in the Barberton greenstone belt (Walsh 1992). However, as confirmed by detailed examination of specimens in petrographic thin sections (Sugitani et al. 2009a) and of extracted specimens by acid maceration (Sugitani et al. 2015a), this morphological type consists of an ellipsoidal to spheroidal central body with a surrounding discoid flange (Figs. 16.6d and 16.11a, b). As described previously, this morphological type includes a variety of subtypes. Variations have been recorded in size, flange width and texture, shape in polar view (circular versus elliptical), and the presence of colonies, which include linked chains, clusters composed solely of lenses and mixtures of lenses and spheroids. Such variations in morphology, together with variations in the mode of occurrence, are best explained either by species diversity (Sugitani et al. 2018) or by life cycle stages.

Morphological variations and modes of occurrence potentially provide information about the life cycle of lenticular microfossils. Their common association in pairs or chains (Fig. 16.6f, g) indicate different stages of binary fission (Sugitani et al. 2015a). On the other hand, colonies composed of lenses and spheroids of vary-

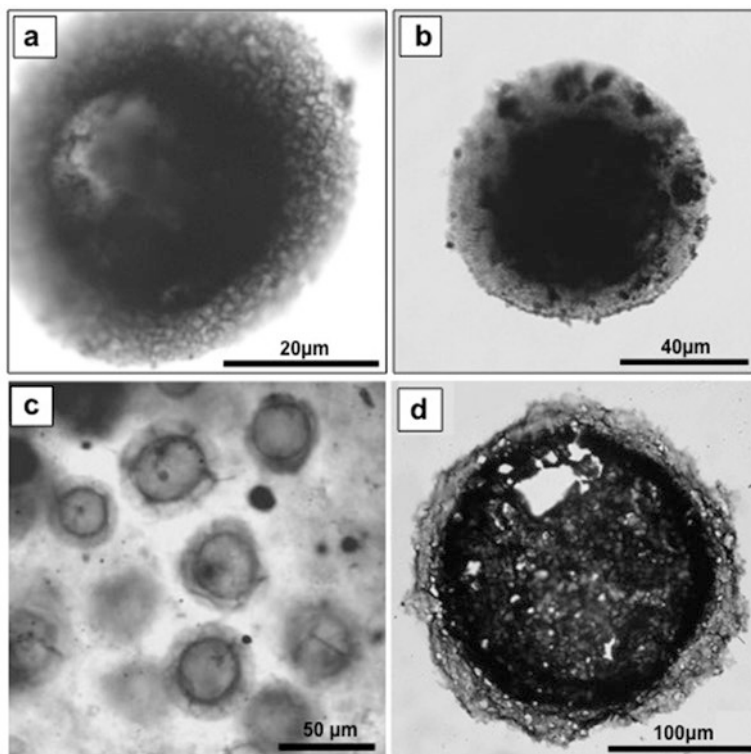


Fig. 16.11 (a, b) Polar view of lenticular microfossils (polar views) extracted by acid maceration, from the Strelley Pool Formation and the Farrel Quartzite, respectively. Specimen in (a) has a reticulate flange, whereas that in (b) has a relatively glassy flange. (c) Early Devonian disphaeromorph (pteromorph) microfossils equivalent to *Pterosperma*, from Rhynie chert (Kustatscher et al. 2014) (with permission from Universal Academy Press). (d) Lenticular microfossil (*Pterospermopsisimorpha pileiformis*) extracted by acid maceration, from the lower Mesoproterozoic Kaltasy Formation, East European Platform (Sergeev et al. 2016) (With permission from Elsevier)

ing sizes (Fig. 16.9b) and lenses containing a single internal sphere (Fig. 16.9c) suggest that lenticular microfossils represent an endospore-bearing resting-cyst stage (Sugitani et al. 2009b). This interpretation is strengthened by a newly discovered specimen (Fig. 16.9d) that appears to be in the process of expelling an inner object. Furthermore, the discovery of specimens that possibly contain several internal objects (Fig. 16.9e) and of colonies apparently composed of small spheroids, lenses, and possible fragmented lenses (Fig. 16.9a) may further suggest reproduction by multiple fissions. The morphologically closest extant microorganism recognized to date is the mature phycoma of *Pterosperma*, a prasinophyte alga that has a flange-like ala (see Parke et al. 1978; Tappan 1980; van den Hoek et al. 1995). This free-swimming, tiny green alga generally reproduces by binary fission but occasionally loses its flagellum and settles on the bottom. Subsequently, it enlarges into a phycoma with an ala. Within the phycoma, multiple zoospores are produced and are

then released to live as a planktonic stage. As described previously, some lenticular microbes might have had a similar life cycle (see Fig. 16.9a) (Sugitani 2012).

Lenticular microfossils with a flange have also been reported from younger successions. Kustatscher et al. (2014) reported lenticular microfossils with an ala, from the Lower Devonian Rhynie chert and assigned them to the phycoma stage of *Pterosperma* (Fig. 16.11c). Spheroid acritarchs with a flange are fairly common in Meso- to Neoproterozoic successions (e.g., Samuelsson 1997; Samuelsson et al. 1999; Vorob'eva et al. 2015; Sergeev et al. 2016) (Fig. 16.11d), although their biological affinity remains to be not confirmed. No detailed comparison of Archean lenticular microfossils with younger morphological equivalents has yet been undertaken, but this could be an effective approach for placing constraints on the possible biological affinities of Archean lenticular microfossils.

Finally, it should be remembered that morphologically similar lenticular microfossils have been discovered in several stratigraphic units ranging in age from 3.0 to 3.4 Ga successions in both Australia and South Africa, so they are more widespread both chronologically and stratigraphically than would be expected. This tends to rule out their abiogenic origin produced by some peculiar physical parameters. They were most probably autotrophic and planktonic based on their carbon isotopic signatures, their mode of occurrence in host rocks, and their morphology (House et al. 2013; Oehler et al. 2017; Kozawa et al. 2018). Moreover, the organisms appear to have had a robust and recalcitrant wall that may have enabled them to have survived harsh environment, including e.g., repeated asteroid impacts (e.g., Lowe et al. 2003; Walsh 1992). It is possible that lenticular microfossils may represent one of the most successful groups in the early history of life on Earth (House et al. 2013; Oehler et al. 2017).

16.6 Conclusions

This review of representative Paleo- and Mesoarchean (>3.0 Ga) microfossils of various morphologies from the Pilbara and Kaapvaal cratons has identified rods, spheroids, lenses, and filaments (and their subtypes). Some spheroidal and lenticular microfossils are significantly large (from 20 μm up to 300 μm), and they had acid-resistant robust organic walls, allowing them to be extracted by acid maceration. Host rocks represent a variety of depositional environments ranging from terrestrial to deep marine, although the majority were deposited in shallow to subaerial environments suitable for photoautotrophs. The possibility that such advanced organisms might have been present is implied by the morphological similarity of some of the small, colonial spheroid microfossils to extant coccoid cyanobacteria, such as *Microcystis*, and by independent geochemical and phylogenetic studies. If this proves to be the case, we can speculate that eukaryotic microorganisms might have emerged in the Archean, possibly in close association with cyanobacteria, although eukaryotes are not commonly thought to have had evolved before the Great Oxidation Event (around 2.45 Ga) (e.g., Javaux et al. 2003; Knoll et al. 2006;

Knoll 2014), and Paleo- and Mesoarchean large microfossils are not necessarily directly related phylogenetically to extant or even Proterozoic eukaryotes. It is possible that the Archean large microfossils, particularly lenticular ones, represent extinct taxa. Nonetheless comparisons with younger equivalents would be an effective approach to determining what the biological affinities were for large microbes in the Paleo- and Mesoarchean. Elucidating the biotic diversity of these microfossils also will provide new insights into the nature of early evolution and diversification of life and how it related to the evolution of the Earth's surface environment. Results from these challenging and innovative studies should also contribute to the search for evidence of life on other planets.

Acknowledgments Financial support from the Japanese Society for the Promotion of Science (Grants-in-aid Nos 22340149 and 24654162) is gratefully acknowledged. I sincerely thank to Kathleen Grey for her constructive review and editing. Tsutomu Nagaoka and Natsuko Takagi are also acknowledged for their assistance for preparation of thin sections.

References

- Allwood AC, Walter MR, Kamber BS, Marshall CP, Burch IW (2006) Stromatolite reef from the Early Archaean era of Australia. *Nature* 441:714–718
- Allwood AC, Burch I, Walter MR (2007) Stratigraphy and facies of the 3.43 Ga Strelley Pool Chert in the Southwestern North Pole Dome, Pilbara Craton, Western Australia. *Geol Sur West Aust Rec* 2007/11
- Allwood AC, Kamber BS, Walter MR, Burch IW, Kanik I (2010) Trace elements record depositional history of an Early Archean stromatolitic carbonate platform. *Chem Geol* 270:148–163
- Alterman W, Kaźmierczak J (2003) Archean microfossils: a reappraisal of early life on earth. *Res Microbiol* 154:611–617
- Angert ER (2005) Alternatives to binary fission in bacteria. *Nat Rev Microbiol* 3:214–224
- Awramik SM, Grey K (2005) Stromatolites: biogenicity, biosignatures, and bioconfusion. In: *Proceedings of SPIE 5906, Astrobiology and Planetart Missions, 59060P*. <https://doi.org/10.1117/12.625556>
- Awramik SM, Schopf JW, Walter MR (1983) Filamentous fossil bacteria from the Archean of Western Australia. *Precambrian Res* 20:357–374
- Awramik SM, Schopf JW, Walter MR (1988) Carbonaceous filaments from North Pole, Western Australia: are they fossil bacteria in Archean stromatolites? A discussion. *Precambrian Res* 39:303–309
- Barghoorn ES, Schopf JW (1966) Microorganisms three billion years old from the Precambrian of South Africa. *Science* 152:758–763
- Barghoorn ES, Tyler SA (1965) Microorganisms from the Gunflint Chert. *Science* 147:563–577
- Brasier MD, Green OR, Jephcoat AP, Kleppe AK, Van Kranendonk MJ, Lindsay JF, Steele A, Grassineau NV (2002) Questioning the evidence for earth's oldest fossils. *Nature* 416:76–81
- Brasier MD, Green OR, Lindsay JF, McLoughlin N, Steele A, Stoakes C (2005) Critical testing of earth's oldest putative fossil assemblage from the ~3.5 Ga Apex chert, Chinaman Creek, Western Australia. *Precambrian Res* 140:55–102
- Brasier M, McLoughlin N, Green O, Wacey D (2006) A fresh look at the fossil evidence for early Archaean cellular life. *Philos Trans R Soc B* 361:887–902
- Brooks J, Muir MD, Shaw G (1973) Chemistry and morphology of Precambrian microorganisms. *Nature* 244:215–217

- Buick R (1984) Carbonaceous filaments from North Pole, Western Australia: are they fossil bacteria in Archaean stromatolites? *Precambrian Res* 24:157–172
- Buick R (1988) Carbonaceous filaments from North Pole, Western Australia: are they fossil bacteria in Archaean stromatolites? A reply. *Precambrian Res* 39:311–317
- Buick (1990) Microfossil recognition in Archean rocks: an appraisal of spheroids and filaments from a 3500 m.y. old chert-barite unit at North Pole, Western Australia. *Palaios* 5:441–459
- Buick R (2008) When did oxygenic photosynthesis evolve? *Philos Trans R Soc B* 363:2731–2743
- Buick R, Dunlop JSR (1990) Evaporitic sediments of early Archaean age from the Warrawoona Group, North Pole, Western Australia. *Sedimentol* 37:247–277
- Cloud PE Jr (1965) Significance of the Gunflint (Precambrian) microflora. *Science* 148:27–35
- Crowe SA, Døssing LN, Beukes NJ, Bau M, Kruger SJ, Frei R, Canfield DE (2013) Atmospheric oxygenation three billion years ago. *Nature* 501:535–538
- Czaja AD, Beukes NJ, Osterhout JT (2016) Sulfur-oxidizing bacteria prior to the great oxidation event from the 2.52 Ga Gamohaan formation of South Africa. *Geology* 44:983–986
- DiMarco MJ, Lowe DR (1989) Shallow-water volcanoclastic deposition in the early Archean Panorama Formation, Warrawoona Group, eastern Pilbara Block, Western Australia. *Sediment Geol* 64:43–63
- 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* 8:15263
- Dodd MS, Papineau D, Grenne T, Slack JF, Rittner M, Pirajno F, O’Neil J, Little CTS (2017) Evidence for early life in Earth’s oldest hydrothermal vent precipitates. *Nature* 543:60–65
- Duck LJ, Glikson M, Golding SD, Webb RE (2007) Microbial remains and other carbonaceous forms from the 3.24 Ga Sulphur Springs black smoker deposit, Western Australia. *Precambrian Res* 154:205–220
- Dunlop JSR, Milne VA, Groves DI, Muir MD (1978) A new microfossil assemblage from the Archaean of Western Australia. *Nature* 274:676–678
- Engel AEJ, Nagy B, Nagy LA, Engel CG, Kremp GOW, Drew CM (1968) Alga-like forms in Onverwacht series, South Africa: oldest recognized lifelike forms on Earth. *Science* 161:1005–1008
- Glikson M, Duck LJ, Golding SD, Hofmann A, Bolhar R, Webb R, Baiano JCF, Sly LI (2008) Microbial remains in some earliest earth rocks: comparison with a potential modern analogue. *Precambrian Res* 164:187–200
- Grey K (1999) A modified palynological preparation technique for the extraction of large Neoproterozoic acanthomorph acritarchs and other acid-insoluble microfossils. *Geol Surv West Aust, Rec* 1999/10, 23 p
- Grey K, Sugitani K (2009) Palynology of Archean microfossils (c. 3.0 Ga) from the Mount Grant area, Pilbara Craton, Western Australia: further evidence of biogenicity. *Precambrian Res* 173:60–69
- Grey K, Roberts FI, Freeman MJ, Hickman AH, Van Kranendonk MJ, Bevan AWR (2010) Management plan for state geoheritage reserves. *Geol Surv West Aust, Rec* 2010/13, 23p
- Heubeck C, Engelhardt J, Byerly GR, Zeh A, Sell B, Lubert T, Lowe DR (2013) Timing of deposition and deformation of the Moodies Group (Barberton Greenstone Belt, South Africa): very-high-resolution of Archaean surface processes. *Precambrian Res* 231:236–262
- Hickman AH (2008) Regional review of the 3426–3350 Ma Strelley Pool Formation, Pilbara Craton, Western Australia. *Geol Surv West Aust, Rec* 2008/15
- Hickman AH (2012) Review of the Pilbara Craton and Fortescue Basin, Western Australia: crustal evolution providing environments for early life. *Island Arc* 21:1–31
- Hickman AH, Van Kranendonk MJ (2008) Archean crustal evolution and mineralization of the northern Pilbara Craton – a field guide. *Geol Surv West Aust, Rec* 2008/13
- Hickman AH, Van Kranendonk MJ, Grey K (2011) State geoheritage reserve R50149 (Trendall Reserve), North Pole, Pilbara Craton, Western Australia – geology and evidence for early Archean life. *Geol Surv West Aust Rec* 2011/10, pp 17–18

- Hoashi M, Bevacqua DC, Otake T, Watanabe Y, Hickman AH, Utsunomiya S, Ohmoto H (2009) Primary haematite formation in an oxygenated sea 3.46 billion years ago. *Nat Geosci* 2:301–306
- Hofmann HJ (2004) Archean microfossils and abiomorphs. *Astrobiology* 4:135–136
- Hofmann A, Harris C (2008) Silica alteration zones in the Barberton greenstone belt: a window into subseafloor processes 3.5–3.3 Ga ago. *Chem Geol* 257:221–239
- Hofmann HJ, Grey K, Hickman AH, Thorpe RI (1999) Origin of 3.45 Ga coniform stromatolites in Warrawoona Group, Western Australia. *Geol Soc Am Bull* 111:1256–1262
- Holland HD (2006) The oxygenation of the atmosphere and oceans. *Philos Trans R Soc Lond B* 361:903–915
- Homann M, Heubeck C, Bontognali TRR, Bouvier A-S, Baumgartner LP, Airo A (2016) Evidence for cavity-dwelling microbial life in 3.22 tidal deposits. *Geology* 4:51–54
- House CH, Oehler DZ, Sugitani K, Mimura K (2013) Carbon isotopic analyses of ca. 3.0 Ga microstructures imply planktonic autotrophs inhabited Earth's early oceans. *Geology* 41:651–654
- Isozaki Y, Kabashima T, Ueno Y, Kitajima K, Maruyama S, Kato Y, Terabayashi M (1997) Early Archean mid-oceanic ridge rocks and early life in the Pilbara Craton, W. Australia. *EOS* 78:399
- Javaux EJ, Knoll AH, Walter MR (2003) Recognizing and interpreting the fossils of early eukaryotes. *Orig Life Evol Biosph* 33:75–94
- Javaux EJ, Marshall CP, Bekker A (2010) Organic-walled microfossils in 3.2-billion-year-old shallow-marine siliciclastic deposits. *Nature* 463:934–938
- Kiyokawa S, Ito T, Ikehara M, Kitajima F (2006) Middle Archean volcano-hydrothermal sequence: bacterial microfossil-bearing 3.2 Ga Dixon Island Formation, coastal Pilbara terrane, Australia. *Geol Soc Am Bull* 118:3–22
- Kiyokawa S, Koge S, Ito T, Ikehara M (2014) An ocean-floor carbonaceous sedimentary sequence in the 3.2-Ga Dixon Island Formation, coastal Pilbara terrane, Western Australia. *Precambrian Res* 255:123–143
- Knoll AH (2014) Paleobiological perspectives on early eukaryotic evolution. *Cold Spring Harb Perspect Biol*. <https://doi.org/10.1101/cshperspect.a016121>
- Knoll AH, Barghoorn ES (1977) Archean microfossils showing cell division from the Swaziland system of South Africa. *Science* 198:396–398
- Knoll AH, Javaux EJ, Hewitt D, Cohen P (2006) Eukaryotic organisms in Proterozoic oceans. *Philos Trans R Soc B* 361:1023–1038
- Kozawa T, Sugitani K, Oehler DZ, House CH, Saito I, Watanabe T, Gotoh T (2018) Early Archean planktonic mode of life: implications from fluid dynamics of lenticular microfossils. *Geobiology*. <https://doi.org/10.1111/gbi.12319>
- Kremer B, Kaźmierczak J (2017) Cellularly preserved microbial fossils from ~3.4 Ga deposits of South Africa: a testimony of early appearance of oxygenic life. *Precambrian Res* 295:117–129
- Kustatscher E, Dotzler N, Taylor TN, Krings M (2014) Microfossils with suggested affinities to the Pyramimonadales (Pyramimonadophyceae Chlorophyta) from the lower Devonian Rhynie chert. *Acta Palaeobotanica* 54:163–171
- Lepot K, Williford KH, Ushikubo T, Sugitani K, Mimura K, Spicuzza MJ, Valley JW (2013) Texture-specific isotopic compositions in 3.4 Gyr old organic matter support selective preservation in cell-like structures. *Geochim Cosmochim Acta* 112:66–86
- Lindsay JF, Brasier MD, McLoughlin N, Green OR, Fogel M, Steele A, Mertzman SA (2005) The problem of deep carbon—an Archean paradox. *Precambrian Res* 143:1–22
- Lowe DR (1999) Petrology and sedimentology of cherts and related silicified sedimentary rocks in the Swaziland Supergroup. *Geol Soc Am Spec Pap* 329:83–114
- Lowe DR, Worrell GF (1999) Sedimentology, mineralogy, and implications of silicified evaporites in the Kromberg Formation, Barberton Greenstone Belt, South Africa. *Geol Soc Am Spec Pap* 329:167–188
- Lowe DR, Byerly GR, Kyte FT, Shukolyukov A, Asaro F, Krull A (2003) Spherule beds 3.47–3.24 billion years old in the Barberton greenstone belt, South Africa: a record of large meteorite impacts and their influence on early crustal and biological evolution. *Astrobiology* 3:7–48

- Lyons TW, Reinhard CT, Planavsky NJ (2014) The rise of oxygen in Earth's early ocean and atmosphere. *Nature* 506:307–315
- Muir MD, Hall DO (1974) Diverse microfossils in Precambrian Onverwacht group rocks of South Africa. *Nature* 252:376–378
- Mukhopadhyay J, Crowely QG, Ghosh S, Ghosh G, Chakrabarti K, Misra B, Heron K, Bose S (2014) Oxygenation of Archean atmosphere: new paleosol constraints from eastern India. *Geology* 42:923–926
- Nagy B, Nagy LA (1969) Early pre-Cambrian Onverwacht microstructures: possibly the oldest fossil on Earth? *Nature* 223:1226–1229
- Nisbet EG, Sleep NH (2001) The habitat and nature of early life. *Nature* 409:1083–1091
- Noffke N, Christian D, Wacey 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
- Nutman AP, Bennett VC, Friend CRL, Van Kranendonk MJ, Chivas AP (2016) Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. *Nature* 537:535–539
- Oehler DZ, Robert F, Walter MR, Sugitani K, Allwood A, Meibom A, Mostefaoui S, Selo M, Thomen A, Gibson EK (2009) NanoSIMS: insights to biogenicity and syngeneity of Archean carbonaceous structures. *Precambrian Res* 173:70–78
- Oehler DZ, Robert F, Walter MR, Sugitani K, Meibom A, Mostefaoui S, Gibson EK (2010) Diversity in the Archean biosphere: new insights from NanoSIMS. *Astrobiology* 10:413–424
- Oehler DZ, Walsh MM, Sugitani K, Liu M-C, House CH (2017) Large and robust lenticular microorganisms on the young earth. *Precambrian Res* 296:112–119
- Parke M, Boalch GT, Jowett R, Harbour DS (1978) The genus *Pterosperma* (Prasinophyceae): species with a single equatorial ala. *J Mar Biol Assoc UK* 58:239–276
- Pflug HD (1967) Structured organic remains from the Fig Tree Series (Precambrian) of the Barberton Mountain Land (South Africa). *Rev Palaeobot Palynol* 5:9–29
- Phoenister V, Konhauser KO, Adams DG, Bottrell SH (2001) Role of biomineralization as an ultraviolet shield: implications for Archean life. *Geology* 29:823–826
- Planavsky NJ, Asael D, Hofmann A, Reinhard CT, Lalonde SV, Knudsen A, Wang X, Ossa-Ossa F, Pecoits E, Smith AJB, Beukes NJ, Bekker A, Johnson TM, Konhauser KO, Lyons TW, Rouxel OJ (2014) Evidence for oxygenic photosynthesis half a billion years before the Great Oxidation Event. *Nat Geosci* 7:283–286
- Rasmussen B (2000) Filamentous microfossils in a 3,235-million-year-old volcanogenic massive sulphide deposit. *Nature* 405:676–679
- Retallack GJ, Krinsley DH, Fischer R, Razink JJ, Langworthy KA (2016) Archean coastal-plain paleosols and life on land. *Gondwana Res* 40:1–20
- Rosing MT, Frei R (2004) U-rich Archean sea-floor sediments from Greenland—indications of > 3700 Ma oxygenic photosynthesis. *Earth Planet Sci Lett* 217:237–244
- Salman V, Amann R, Girmth A-C, Polerecky L, Bailey JV, Högslund S, Jessen G, Pantoja S, Schult-Vogt HN (2011) A single-cell sequencing approach to the classification of large, vacuolated sulfur bacteria. *Syst Appl Microbiol* 34:243–259
- Salman V, Bailey JV, Teske A (2013) Phylogenetic and morphologic complexity of giant sulphur bacteria. *Antonie Van Leeuwenhoek* 104:169–189
- Samuelsson J (1997) Biostratigraphy and paleobiology of Early Neoproterozoic strata of the Kola Peninsula, Northwest Russia. *Nor Geol Tidsskr* 77:165–192
- Samuelsson J, Dawes PR, Vidal G (1999) Organic-walled microfossils from the Proterozoic Thule Supergroup, Northwest Greenland. *Precambrian Res* 96:1–23
- Schirmermeister BE, Sanchez-Baracaldo P, Wacey D (2016) Cyanobacterial evolution during the Precambrian. *Int J Astrobiol* 15:187–204
- Schopf JW (1976) Are the oldest 'fossils', fossils? *Orig Life* 7:19–36
- Schopf JW (1993) Microfossils of the Early Archean Apex chert: new evidence of the antiquity of life. *Science* 260:640–646
- Schopf JW (2006) Fossil evidence of Archean life. *Philos Trans R Soc B* 361:869–885

- Schopf JW, Barghoorn ES (1967) Alga-like fossils from the Early Precambrian of South Africa. *Science* 156:508–512
- Schopf JW, Packer BM (1987) Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. *Science* 237:70–73
- Schopf JW, Walter MR (1983) Archean microfossils: new evidence of ancient microbes. In: Schopf JW (ed) *Earth's earliest biosphere, its origin and evolution*. Princeton Univ. Press, Princeton, pp 214–239
- Schopf JW, Hayes JM, Walter MR (1983) Evolution of Earth's earliest ecosystem: recent progress and unsolved problems. In: Schopf JW (ed) *Earth's earliest biosphere. Its origin and evolution*. Princeton University Press, Princeton, pp 361–384
- Schopf JW, Kudryavtsev AB, Sugitani K, Walter MR (2010) Precambrian microbe-like pseudofossils: a promising solution to the problem. *Precambrian Res* 179:191–205
- Schopf JW, Kudryavtsev AB, Walter MR, Van Kranendonk MJ, Williford KH, Kozdon R, Valley JW, Gallardo VA, Espinoza C, Flannery DT (2015) Sulfur-cycling fossil bacteria from the 1.8-Ga Duck Creek Formation provide promising evidence of evolution's null hypothesis. *Proc Natl Acad Sci* 112:2087–2092
- Schopf JW, Kudryavtsev AB, Osterhout JT, Williford KH, Kitajima K, Valley JW, Sugitani K (2017) An anaerobic ~3400 Ma shallow-water microbial consortium: presumptive evidence of Earth's Paleoproterozoic anoxic atmosphere. *Precambrian Res* 299:309–318
- Schopf JW, Kitajima K, Spicuzza MJ, Kudryavtsev AB, Valley JW (2018) SIMS analyses of the oldest known assemblage of microfossils document their taxon-correlated carbon isotope compositions. *Proc Natl Acad Sci USA* 115:53–58
- Schulz HN, Brinkhoff T, Ferdelman TG, Mariné MH, Teske A, Jørgensen BB (1999) Dense populations of a giant sulfur bacterium in Namibian shelf sediments. *Science* 284:493–495
- Sergeev VN, Knoll AH, Vorob'eva NG, Sergeeva ND (2016) Microfossils from the lower Mesoproterozoic Kaltasy Formation, East European platform. *Precambrian Res* 278:87–107
- Siever R (1992) The silica cycle in the Precambrian. *Geochim Cosmochim Acta* 56:3265–3272
- Stefurak EJT, Lowe DR, Zenter D, Fischer WW (2014) Primary silica granules – a new mode of Paleoproterozoic sedimentation. *Geology* 42:283–286
- Sugahara H, Sugitani K, Mimura K, Yamashita F, Yamamoto K (2010) A systematic rare-earth elements and yttrium study of Archean cherts at the Mount Goldsworthy greenstone belt in the Pilbara Craton: implications for the origin of microfossil-bearing black cherts. *Precambrian Res* 177:73–87
- Sugitani K (2012) Life cycle and taxonomy of Archean flanged microfossils from the Pilbara Craton, Western Australia. 34th International Geological Congress (IGC): AUSTRALIA 2012, 17.3#257
- Sugitani K, Mimura K, Suzuki K, Nagamine K, Sugisaki R (2003) Stratigraphy and sedimentary petrology of an Archean volcanic-sedimentary succession at Mt. Goldsworthy in the Pilbara Block, Western Australia: implications of evaporite (nahcolite) and barite deposition. *Precambrian Res* 120:55–79
- Sugitani K, Grey K, Allwood AC, Nagaoka T, Mimura K, Mimura M, Marshall CP, Van Kranendonk MJ, Walter MR (2007) Diverse microstructures from Archean chert from the Mount Goldsworthy – Mount Grant area, Pilbara Craton, Western Australia: microfossils, dubiomicrofossils, or pseudofossils? *Precambrian Res* 158:228–262
- Sugitani K, Grey K, Nagaoka T, Mimura K (2009a) Three-dimensional morphological and textural complexity of Archean putative microfossils from the northeastern Pilbara Craton: indications of biogenicity of large (>15µm) spheroidal and spindle-like structures. *Astrobiology* 9:603–615
- Sugitani K, Grey K, Nagaoka T, Mimura K, Walter MR (2009b) Taxonomy and biogenicity of Archean spheroidal microfossils (ca. 3.0 Ga) from the Mount Goldsworthy-Mount Grant area in the northeastern Pilbara Craton, Western Australia. *Precambrian Res* 173:50–59
- Sugitani K, Lept K, Nagaoka T, Mimura K, Van Kranendonk M, Oehler DZ, Walter MR (2010) Biogenicity of morphologically diverse carbonaceous microstructures from the ca. 3400 Ma Strelley Pool Formation, in the Pilbara Craton, Western Australia. *Astrobiology* 10:899–920

- Sugitani K, Mimura K, Nagaoka T, Lepot K, Takeuchi M (2013) Microfossil assemblage from the 3400Ma Strelley Pool Formation in the Pilbara Craton, Western Australia: results from a new locality. *Precambrian Res* 226:59–74
- Sugitani K, Mimura K, Takeuchi M, Lepot K, Ito S, Javaux EJ (2015a) Early evolution of large micro-organisms with cytological complexity revealed by microanalyses of 3.4 Ga organic-walled microfossils. *Geobiology* 13:507–521
- Sugitani K, Mimura K, Takeuchi M, Yamaguchi T, Suzuki K, Senda R, Asahara Y, Wallis S, Van Kranendonk MJ (2015b) A Paleoproterozoic coastal hydrothermal field inhabited by diverse microbial communities: the Strelley Pool Formation, Pilbara Craton, Western Australia. *Geobiology* 13:522–545
- Sugitani K, Van Kranendonk MJ, Oehler DZ, House CH, Walter MR (2017) Comment: Archean coastal-plain paleosols and life on land. *Gondwana Res* 44:265–269
- Sugitani K, Kohama T, Mimura K, Takeuchi M, Senda R, Morimoto H (2018) Speciation of Paleoproterozoic life demonstrated by analysis of the morphological variation of lenticular microfossils, from the Pilbara Craton of Western Australia. *Astrobiology* 18:1057–1070
- Tappan H (1980) The paleobiology of plant protists. W.H. Freeman and Co., San Francisco, 1028 p
- Tice MM, Lowe DR (2004) Photosynthetic microbial mats in the 3,416-Myr-old-ocean. *Nature* 431:549–552
- Tice MM, Lowe DR (2006a) The origin of carbonaceous matter in pre-3.0 Ga greenstone terranes: a review and new evidence from the 3.42 Ga Buck Reef Chert. *Earth Sci Rev* 76:259–300
- Tice MM, Lowe DR (2006b) Hydrogen-based carbon fixation in the earliest known photosynthetic organisms. *Geology* 34:37–40
- Ueno Y, Isozaki Y, Yurimoto H, Maruyama S (2001a) Carbon isotopic signatures of individual Archean microfossils (?) from Western Australia. *Int Geol Rev* 43:196–212
- Ueno Y, Maruyama S, Isozaki Y, Yurimoto H (2001b) Early Archean (ca. 3.5 Ga) microfossils and ¹³C-depleted carbonaceous matter in the North Pole area, Western Australia: field occurrence and geochemistry. In: Nakashima S et al (eds) *Geochemistry and the origin of life*. Universal Academy Press, Tokyo, pp 203–236
- Ueno Y, Yamada K, Yoshida N, Maruyama S, Isozaki Y (2006) Evidence from fluid inclusions for microbial methanogenesis in the early Archean era. *Nature* 440(7083):516–519
- van den Hoek C, Mann DG, Jahns HM (1995) *Algae: an introduction to phycology*. Cambridge University Press, Cambridge, UK 625 p
- Van Kranendonk MJ (2006) Volcanic degassing, hydrothermal circulation and the flourishing of early life on earth: a review of the evidence from c. 3490–3240 Ma rocks of the Pilbara Supergroup, Pilbara Craton, Western Australia. *Earth Sci Rev* 74:197–240
- Van Kranendonk MJ (2007) A review of the evidence for putative Paleoproterozoic life in the Pilbara Craton, Western Australia. In: Van Kranendonk MJ et al (eds) *Earth's oldest rocks*. Elsevier, Amsterdam, pp 855–877
- Van Kranendonk MJ (2011) Morphology as an indicator of biogenicity for 3.5–3.2 Ga fossil stromatolites from the Pilbara Craton, Western Australia. In: Reitner J, Quéric N-V, Arp G (eds) *Advances in stromatolite geobiology, Lecture Notes in Earth Sciences, vol 131*. Springer, Cham, pp 537–554
- Van Kranendonk MJ, Hickman AH, Smithies RH, Nelson DN, Pike G (2002) Geology and tectonic evolution of the Archean North Pilbara terrain, Pilbara Craton, Western Australia. *Econ Geol* 97:695–732
- Van Kranendonk MJ, Hickman AH, Smithies RH, Williams IR, Bagas L, Farrell TR (2006) Revised lithostratigraphy of Archean supracrustal and intrusive rocks in the northern Pilbara Craton, Western Australia. *West Aust Geol Surv, Rec* 2006/15
- Van Kranendonk MJ, Philippot P, Lepot K, Bodorkos S, Pirajno F (2008) Geological setting of earth's oldest fossils in the ca. 3.5 Ga Dresser formation, Pilbara Craton, Western Australia. *Precambrian Res* 167:93–124

- Vearncombe S, Barely ME, Groves DI, McNaughton NJ, Mikucki EJ, Vearncombe JR (1995) 3.26 Ga black smoker-type mineralization in the Strelley Pool Belt, Pilbara Craton, Western Australia. *J Geol Soc Lond* 152:587–590
- Vorob'eva NG, Sergeev VN, Petrov PY (2015) Kotuikan formation assemblage: a diverse organic-walled microbiota in the Mesoproterozoic Anabar succession, northern Siberia. *Precambrian Res* 256:201–222
- Wacey D (2009) *Early life on earth: a practical guide*. Springer, Heidelberg
- Wacey D (2012) Earliest evidence for life on earth: an Australian perspective. *Aust J Earth Sci* 59:153–166
- Wacey D, Kilburn MR, Saunders M, Cliff J, Brasier MD (2011) Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nat Geosci* 4:698–702
- Wacey D, Noffke N, Saunders M, Guagliardo P, Pyle DM (2018a) Volcanogenic pseudo-fossils from the ~3.48 Ga Dresser Formation, Pilbara, Western Australia. *Astrobiology* 18:539–555
- Wacey D, Saunders M, Kong C (2018b) Remarkably preserved tephra from the 3430 Ma Strelley Pool Formation, Western Australia: implications from the interpretation of Precambrian microfossils. *Earth Planet Sci Lett* 487:33–43
- Walsh MM (1992) Microfossils and possible microfossils from Early Archean Onverwacht Group, Barberton Mountain Land, South Africa. *Precambrian Res* 54:271–293
- Walsh MM, Lowe DR (1985) Filamentous microfossils from the 3,500-Myr-old Onverwacht Group, Barberton Mountain Land, South Africa. *Nature* 314:530–532
- Walter MR, Bauld J, Brock TD (1976) Microbiology and morphogenesis of columnar stromatolites (*Conophyton*, *Vacerrilla*) from hot springs in Yellowstone National Park. In: Walter MR (ed) *Stromatolites, Developments in Sedimentology* 20. Elsevier, Amsterdam, pp 273–310
- Waterbury JB, Stanier RY (1978) Patterns of growth and development in pleurocapsalean cyanobacteria. *Microbiol Rev* 42:2–44
- Westall F, de Wit MJ, Dann J, van der Gaast S, de 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
- Westall F, de Vries ST, Nijman W, Rouchon V, Orberger B, Pearson V, Watson J, Verchovsky A, Wright I, Rouzaud J-N, Marchesini D, Severine A (2006) The 3.446 Ga “Kitty’s Gap Chert”, an early Archean microbial ecosystem. *Geol Soc Am Bull, Spec Pap* 405:105–131
- Westall F, Foucher F, Cavalazzi B, de Vries ST, Nijman W, Pearson V, Watson J, Verchovsky A, Wright I, Rouzaud J-N, Marchesini D, Anne S (2011) Volcaniclastic habitats for early life on Earth and Mars: a case study from ~3.5 Ga-old rocks from the Pilbara, Australia. *Planet Space Sci* 310:1093–1106
- Witze A (2016) Claims of earth’s oldest fossils tantalize researchers. *Nature*. <https://doi.org/10.1038/nature.2016.20506>
- Wolf ET, Toon OB (2010) Fractal organic haze provided an ultraviolet shield for early earth. *Science* 328:1266–1268

Chapter 17

Great Oxidation Event and Snowball Earth



Eiichi Tajika and Mariko Harada

Abstract The atmosphere of early Earth contained little molecular oxygen. A significant increase in oxygen occurred ca. 2.4–2.0 billion years ago in what is called the Great Oxidation Event (GOE). A large positive excursion in carbon isotopic composition in sedimentary carbonates is known to have occurred 2.2–2.0 billion years ago (the Lomagundi-Jatuli event), which provides evidence for an enhanced rate of organic carbon burial, i.e., enhanced net production of oxygen. The Proterozoic snowball Earth event (global glaciation) occurred 2.3–2.2 billion years ago, roughly coinciding with the GOE. Thus, a causal relationship between the GOE and the snowball Earth event has been suggested. The snowball Earth event could have been triggered by an increase in oxygen in the atmosphere because it would have resulted in a significant reduction of atmospheric methane level, thereby reducing the greenhouse effect of the atmosphere and causing global glaciation. On the other hand, termination of the snowball Earth event may have triggered the production of a large amount of oxygen because the extremely hot climate ($\sim 60^\circ\text{C}$) immediately after the termination of the snowball Earth event must have significantly increased the supply of phosphate to the oceans, resulting in large-scale blooms of cyanobacteria, which could have produced large amounts of oxygen. The postglacial transition of atmospheric oxygen levels may have promoted an ecological shift and biological innovations for oxygen-dependent life.

Keywords Great oxidation event · Snowball Earth · Oxygen · Cyanobacteria

E. Tajika (✉)

Department of Earth and Planetary Science, Graduate School of Science,
The University of Tokyo, Bunkyo-ku, Tokyo, Japan
e-mail: tajika@eps.s.u-tokyo.ac.jp

M. Harada

Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: harada.mariko.gn@u.tsukuba.ac.jp

17.1 Introduction

Molecular oxygen (O_2) makes up 20.9% by volume of the atmosphere today. It was, however, only $<10^{-12}$ times the present atmospheric level (PAL) on early Earth (Kasting 1993). Cyanobacteria are considered to have produced O_2 and changed the redox (reduction-oxidation) condition of the surface environment of Earth. The transition from an anaerobic to an aerobic environment must have, in turn, significantly influenced life and its evolution. This is one of the central issues involving the coevolution of Earth and life.

Evolution of atmospheric O_2 level has been a matter of debate for many decades (e.g., Berkner and Marshall 1965; Cloud 1972; Holland 1984; Kasting 1993; Lyons et al. 2014). Based on the geological record and geochemical analyses, it has been recognized that there were two events of the significant rise in the atmospheric O_2 level in the history of the Earth: one is called the “Great Oxidation Event” (GOE)—which occurred in the Paleoproterozoic between ca. 2.4 and 2.0 billion years ago (Ga) (Holland 2002)—and the other is called the “Neoproterozoic Oxygenation Event” (NOE), which occurred during the Neoproterozoic, ca. 800–600 million years ago (Ma) (Shields-Zhou and Och 2011) (Fig. 17.1). Atmospheric O_2 levels rose from $<10^{-5}$ to 0.01–0.001 PAL during the GOE, resulting in the greatest environmental change in the history of the Earth. The reason why the atmospheric O_2 levels rose at that time remains to be ascertained, although there are many hypotheses proposed until now (see Sect. 17.2).

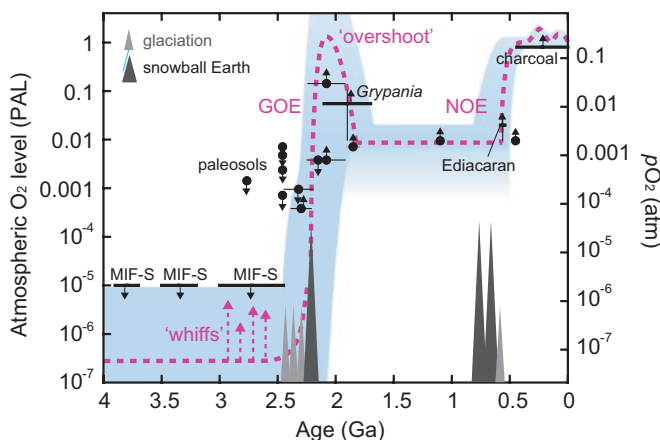


Fig. 17.1 The history of O_2 in the Earth’s atmosphere (Harada et al. 2015). The atmospheric O_2 levels (in PAL; i.e., normalized to the present atmospheric level) would have risen mainly in two geological periods in the Earth history: the Great Oxidation Event (GOE) at 2.4–2.0 Ga and the Neoproterozoic Oxygenation Event (NOE) at 0.8–0.6 Ga. Overshoot of O_2 occurred at 2.2–2.1 Ga and whiffs of O_2 occurred before the GOE. Glaciations in the Proterozoic (triangles)—especially snowball Earth glaciations (black triangles)—coincided with increases in the O_2 level during both GOE and NOE. Arrows with filled and open circles and bars are constraints from geochemical studies (e.g., Farquhar et al. 2007; Goto et al. 2013; Klemm 2000; Pavlov and Kasting 2002)

Another type of environmental change that may have had a significant impact on the evolution of life would be global glaciations, i.e., “snowball Earth” events (e.g., Kirschvink 1992; Hoffman et al. 1998; Hoffman and Schrag 2002). During the snowball Earth event, the surface temperature decreased to $-40\text{ }^{\circ}\text{C}$, and the entire surface of Earth was covered with ice. Life would have faced serious crisis because liquid water is essential for life. The snowball Earth event is known to have occurred at least thrice in Earth’s history: 2.3–2.2 Ga, 720–663 Ma, and 639–635 Ma (Fig. 17.1) (e.g., Kirschvink et al. 2000; Shields-Zhou et al. 2016).

It is interesting that spikes in atmospheric O_2 levels roughly coincided with snowball Earth events, both in the Paleoproterozoic and in the Neoproterozoic (Fig. 17.1). Hence, there might have been a causal relationship between them. In particular, the Paleoproterozoic snowball Earth event could have been caused by a rise in O_2 (e.g., Kirschvink et al. 2000; Pavlov et al. 2000, 2003; Kasting et al. 2001; Kasting 2005; Kopp et al. 2005; Claire et al. 2006). Alternately, a significant rise in the atmospheric O_2 could have been triggered by the Paleoproterozoic snowball Earth event (e.g., Kirschvink et al. 2000; Harada et al. 2015).

In this chapter, recent progress of studies on the rise in atmospheric O_2 levels during the Paleoproterozoic (i.e., GOE) and the Paleoproterozoic snowball Earth events are reviewed and their possible relationship discussed.

17.2 Great Oxidation Event

Atmospheric O_2 levels were very low before 2.45 Ga, as inferred from the formation of deposits of detrital pyrite and uraninite (e.g., Holland 1984; Rasmussen and Buick 1999). The mass-independent fractionation of sulfur isotopes in sulfide and sulfate minerals (MIF-S) provides further strong evidence for low O_2 levels ($<10^{-5}$ PAL) before 2.45 Ga (e.g., Farquhar et al. 2000; Pavlov and Kasting 2002). However, the enrichment of redox-sensitive trace metals such as molybdenum and rhenium in sedimentary rocks before 2.5 Ga suggests episodes of small increases in atmospheric O_2 levels in the environment in the late Archean, called “whiffs” of oxygen (e.g., Anbar et al. 2007). The mass-independent fractionation of sulfur isotopes disappeared by 2.32 Ga, providing evidence for the rise in O_2 for the first time between 2.32 Ga and 2.45 Ga (Bekker et al. 2004).

Atmospheric O_2 seems to have increased to a significant level by 2.2 Ga, inferred from the worldwide appearance of red beds (oxidized subaerial deposits formed via the oxidative weathering of land soils) (e.g., Chandler 1980; Rye and Holland 1998; Bekker and Holland 2012). Carbon isotope records in sedimentary carbonates deposited between 2.22 and 2.06 Ga also show large positive carbon isotope excursions, up to $+16\text{‰}$ in $\delta^{13}\text{C}$ and even higher (e.g., Karhu and Holland 1996; Martin et al. 2013). The enrichment in ^{13}C suggests large perturbations to the carbon cycle system, causing high rates of organic carbon burial. The increase in the rate of organic carbon burial resulted in the net production of between 12 and 22 times the present atmospheric amount of O_2 (Karhu and Holland, 1996). This

“Lomagundi-Jatuli event” is recognized in a number of sedimentary basins worldwide, and the positive anomaly of carbon isotope data provides direct evidence for the production of a large amount of O₂ in this interval (Karhu and Holland 1996; Bekker et al. 2006; Bekker and Holland 2012).

The deposition of banded iron formations (BIFs) has been considered to associate with the rise in O₂ (e.g., Cloud 1973; Holland 1984; Kasting 1993, 2013). Although many of these formations might have been formed by the activity of anaerobic iron-oxidizing bacteria (Konhauser et al. 2002), the occurrence of large amounts of BIFs just before ca. 2.4 Ga and their disappearance for several hundred million years afterwards are consistent with the rise in O₂ occurring since ca. 2.4 Ga (Isley and Abott 1999).

Lines of geological and geochemical evidence indicate that the Lomagundi-Jatuli event was accompanied by an “oxygen overshoot” (Fig. 17.1). The overshoot is inferred from the worldwide deposition of marine sulfate evaporites, an increase in the ratio of average ferric iron to total iron in shales, and the enrichment of other redox-sensitive trace metals in sedimentary rocks (Bekker et al. 2006; Schroder et al. 2008; Bekker and Holland 2012; Canfield et al. 2013).

The cause for the GOE has been a matter of debate for a long time. Cyanobacteria are thought to have been responsible for the rise in O₂, while it is not known when cyanobacteria emerged (e.g., Brocks et al. 1999; Summons et al. 1999). Oxygen could have been produced for several hundred million years before the GOE, which is suggested by the evidence for oxic conditions inferred from studies on redox-sensitive trace metals (e.g., Anbar et al. 2007), molybdenum isotopes (Planavsky et al. 2014), chromium isotopes (Frei et al. 2009; Crowe et al. 2013), and carbon isotopes (Hayes 1983) in sedimentary rocks of the late Archean age (2.5–3.0 Ga). If that were the case, why did O₂ not accumulate in the atmosphere for several hundred million years until ca. 2.4 Ga?

The classical idea for this is that the supply rate of reductants such as H₂ and Fe²⁺ from the Earth’s interior was larger than the supply rate of O₂ produced by cyanobacteria. The balance changed around the GOE: the supply rate of reductants decreased, and/or the supply rate of O₂ increased. The former condition includes (1) decreases in the amount of reducing volcanic gases with time due to oxidation of the upper mantle through the reaction between iron and water followed by the loss of hydrogen to space (Kasting et al. 1993), (2) decreases in reducing volcanic gases due to a gradual shift from submarine to subaerial volcanism around the GOE (Kump and Barley 2007; Gaillard et al. 2011), or (3) decreases in reducing metamorphic gases due to the oxidation of continental crust (Catling et al. 2001; Claire et al. 2006). The latter condition includes increase in the burial rate of organic carbon (which corresponds to increase in the net production rate of O₂) owing to (1) the development of a shallow continental margin where organic carbon is buried efficiently, (2) increase in marine biological productivity because of an increase in the input rate of bio-limiting nutrients into the oceans, and (3) increase in marine biological productivity from an increase in the availability of molybdenum, which is a key component of the nitrogenase enzyme used by cyanobacteria to fix atmospheric nitrogen (Zerkle et al. 2006; Scott et al. 2011).

Another hypothesis to account for the delay in the accumulation of O₂ is that atmospheric O₂ is bistable (Goldblatt et al. 2006). Nonlinearity in the rate of photochemical reaction between O₂ and methane (CH₄) in the atmosphere causes two different stable steady-state O₂ levels. When the atmospheric O₂ increases to the level where an ozone screen forms, rapid transition to an oxidative condition (> ~0.01 PAL) can occur (Goldblatt et al. 2006).

In spite of the hypotheses proposed so far, no consensus has been made on the actual cause for the GOE. For more comprehensive reviews on the GOE and its possible causes, see Kasting (2013), Lyons et al. (2014), Catling and Kasting (2017), and references therein.

17.3 Paleoproterozoic Snowball Earth

Evidence for glacial sediments deposited at low paleolatitude during the Marinoan glaciation (~640 Ma) was reported from the Elatina Formation of South Australia (Embleton and Williams 1986). The low-latitude magnetic direction was confirmed to be of primary origin (Sumner et al. 1987). It is therefore concluded that there were continental ice sheets at sea level near the equator at that time. To interpret the result, Kirschvink (1992) proposed a “snowball Earth” hypothesis, which suggests the Earth to have been globally glaciated at that time.

When the greenhouse effect of the atmosphere decreases for any reason, the climate of the Earth cools, and polar ice expands from mid- to low-latitudes, resulting in the surface of the Earth being covered with ice globally, owing to a large ice-cap instability caused by the ice-albedo feedback mechanism (North et al. 1981). The surface temperature decreases to -40 °C owing to high ice albedo (Fig. 17.2).

The surface temperature, however, increases with time during a snowball stage because CO₂ degassing from the interior of the Earth via volcanism accumulates in the atmosphere, since chemical weathering and photosynthesis do not occur under frozen surface conditions (Kirschvink 1992) (Fig. 17.2). When CO₂ partial pressure became 0.12 bar at solar luminosity 94% of that at present in the Neoproterozoic, or 0.7 bar at solar luminosity 83% of that at present in the Paleoproterozoic, ice melted from the equator to the poles (Caldeira and Kasting 1992; Tajika 2003). After deglaciation, the surface temperature increased to 60 °C (Fig. 17.2) because there was still a large amount of CO₂ in the atmosphere. The surface temperature then returned to normal levels owing to intensive chemical weathering of silicate rocks on the continents followed by the precipitation of carbonates in the oceans.

Evidence for low-latitude glaciation has been also found in the Sturtian (~700 Ma) (Park 1997) and Makganyene glaciations (~2.3 Ga) (Evans et al. 1997). Accordingly, there are at least three snowball Earth events known in Earth’s history.

In the Transvaal Supergroup in the Griqualand West region of South Africa, the uppermost glacial diamictite named the Makganyene Diamictite Formation is overlain with volcanic lavas for which an age of 2.222 ± 0.013 Ga and paleolatitude of

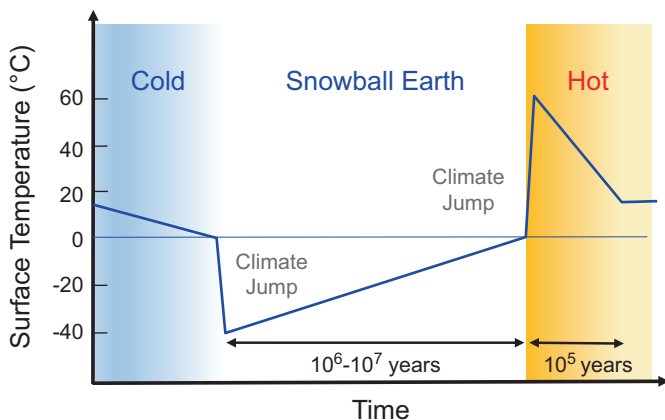


Fig. 17.2 Schematic variation of the average surface temperature during the snowball Earth event. The surface temperature reduced to -40°C when Earth became globally covered with ice, increased gradually owing to accumulation of CO_2 in the atmosphere via volcanism, and then became 60°C immediately after deglaciation because of very high $p\text{CO}_2$ (on the order of 0.1 bar) accumulated in the atmosphere (e.g., Tajika 2003)

$11 \pm 5^{\circ}$ are reported, providing evidence for low-latitude glaciation, i.e., the Paleoproterozoic snowball Earth event (Evans et al. 1997; Kirschvink et al. 2000).

Details of the Makganyene snowball Earth event are not known well, partly because stratigraphic correlations between different sedimentary basins are controversial (e.g., Hilburn et al. 2005; Sekine et al. 2011; Rasmussen et al. 2013). However, stratigraphic features of the Transvaal Supergroup suggest that a significant rise in O_2 occurred just after the deposition of glacial diamictites (Kirschvink et al. 2000). In the next section, a possible relationship between the increase in O_2 and the Makganyene snowball Earth event will be discussed.

17.4 Possible Relationship Between the Snowball Earth and the Great Oxidation Event

The Paleoproterozoic snowball Earth event is suggested to have been triggered by an increase in O_2 in the atmosphere. This is because the rise in O_2 would have resulted in the oxidation of CH_4 , which is considered to have played an important role in sustaining the warm climate in the Archean (Pavlov et al. 2000; Kasting et al. 2001; Kasting 2005; Ozaki et al. 2017). The increase in the atmospheric O_2 level led to immediate decrease in the greenhouse effect due to CH_4 , which could have caused global glaciation (e.g., Pavlov et al. 2000; Kasting et al. 2001; Kirschvink et al. 2002; Kasting 2005; Kopp et al. 2005; Goldblatt et al. 2006; Kirschvink and Kopp 2008; Claire et al. 2006).

It has been proposed that the significant increase in O_2 was caused by the Paleoproterozoic snowball Earth event (e.g., Kirschvink et al. 2000; Harada et al.

2015). In the Transvaal Supergroup, the first large-scale sedimentary manganese (Mn) ore deposits in Earth's history formed immediately above the Makganyene Diamictite Formation (Kirschvink et al. 2000), although the origin of the Paleoproterozoic Mn oxides has been debated (e.g., Gnos et al. 2003; Johnson et al. 2013; Kurzweil et al. 2016). Because Mn has a high oxidation potential, it is suggested that the rise in O_2 level occurred just after the Makganyene snowball Earth event (Kirschvink et al. 2000).

Intense chemical weathering of continents must have occurred in the very hot climate ($\sim 60^\circ C$) just after deglaciation to deliver a large quantity of bio-limiting nutrients such as phosphorus into the oceans. This may have resulted in large-scale blooms of cyanobacteria producing a massive amount of O_2 to trigger the transition from low- to high-stable steady states of the atmospheric O_2 level (Fig. 17.3) (Harada et al. 2015).

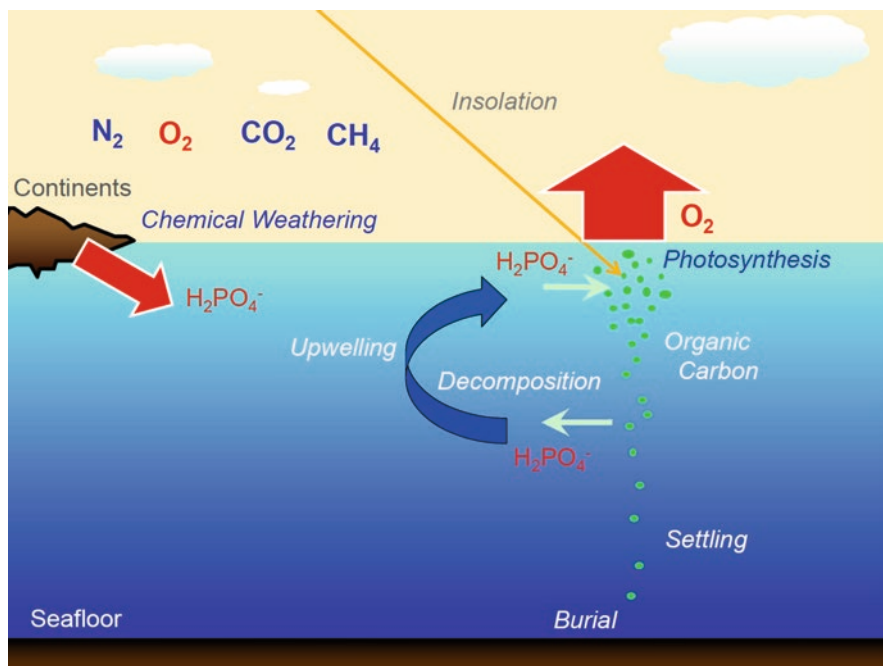


Fig. 17.3 Schematic scenario for the rise in O_2 immediately after the termination of the snowball Earth event (Harada et al. 2015). The rate of chemical weathering of continental rocks was one order of magnitude higher than the present rate under the very hot ($\sim 60^\circ C$) climate condition immediately after deglaciation. A large amount of bio-limiting nutrients including phosphate ($H_2PO_4^-$)—which are necessary for cyanobacteria to photosynthesize—must have been delivered to the oceans. Such an exceptionally large-scale perturbation to the biogeochemical cycle system may have caused significant blooms of cyanobacteria, resulting in net production of vast amounts of O_2 , released to the atmosphere. Therefore, a transition of atmospheric O_2 level from low ($\sim 10^{-6}$ PAL) to high (~ 0.01 PAL) stable steady states must have occurred, with extensive oxygen overshoot. Filled circles represent organic matter produced by photosynthesis in the surface oceans, which are decomposed during settling, and buried in the seafloor sediments (see Harada et al. (2015) for details)

Numerical simulations showed that the O₂ level increased to 1 PAL soon after deglaciation and then decreased to 0.01 PAL over a timescale of the order of 10⁸ years, representing an overshoot of O₂ level. Such a large perturbation does not usually occur, but can occur only after the termination of a snowball Earth event (Harada et al. 2015).

The increase in O₂ with extensive overshoot may have caused global oxygenation of Earth's surface and must have promoted a shift in the marine ecosystem from anaerobes to aerobes. This environmental stress may have driven biological innovations including endosymbiosis, leading to the prosperity of oxygen-dependent complex life.

17.5 Conclusion

During the Paleoproterozoic, the atmospheric O₂ level increased greatly, and the redox conditions in the atmosphere and the surface ocean system changed dramatically. This must have influenced life significantly. Earth was globally glaciated during the Paleoproterozoic, which must have also impacted life and its evolution. Because the Paleoproterozoic snowball Earth event seems to have coincided with the GOE, a causal relationship between the two has been suggested. However, a process independent from glaciation may have triggered the first rise in O₂ at 2.45–2.32 Ga. Further investigations are needed to reveal the detailed behavior of the atmospheric O₂ level and the Earth system during the GOE and also to understand the coevolution of Earth and life in one of the most important periods of Earth's history.

References

- Anbar AD, Duan YT, Lyons TW, Arnold GL, Kendall B, Creaser RA, Kaufman AJ, Gordon GW, Scott C (2007) A whiff of oxygen before the great oxidation event? *Science* 317(5846):1903–1906
- Bekker A, Holland HD (2012) Oxygen overshoot and recovery during the early Paleoproterozoic. *Earth Planet Sci Lett* 317:295–304
- Bekker A, Holland HD, Wang P-L, Rumble D III, Stein HJ, Hannah JL, Coetzee LL, Beukes NJ (2004) Dating the rise of atmospheric oxygen. *Nature* 427(6970):117–120
- Bekker A, Karhu JA, Kaufman AJ (2006) Carbon isotope record for the onset of the Lomagundi carbon isotope excursion in the Great Lakes area, North America. *Precambrian Res* 148:145–180
- Berkner LV, Marshall LC (1965) On the origin and rise of oxygen concentration in the Earth's atmosphere. *J Atmos Sci* 22:225–261
- Brocks JJ, Logan GA, Buick R, Summons RE (1999) Archean molecular fossils and the early rise of eukaryotes. *Science* 285:1033–1036
- Caldeira K, Kasting JF (1992) Susceptibility of the early Earth to irreversible glaciation caused by carbon dioxide clouds. *Nature* 359:226–228

- Canfield DE, Ngombi-Pemba L, Hammarlund EU, Bengtson S, Chaussidon M, Gauthier-Lafaye F, Meunier A, Riboulleau A, Rollion-Bard C, Rouxel O, Asael D, Pierson-Wickmann A-C, El Albani A (2013) Oxygen dynamics in the aftermath of the great oxidation of Earth's atmosphere. *Proc Natl Acad Sci U S A* 110:16736–16741
- Catling DC, Kasting JF (2017) Atmospheric evolution on inhabited and lifeless worlds. Cambridge University Press, Cambridge, UK, 592p
- Catling DC, Zahnle KJ, McKay CP (2001) Biogenic methane, hydrogen escape, and the irreversible oxidation of early Earth. *Science* 293:839–843
- Chandler FW (1980) Proterozoic redbed sequences of Canada. *Can Geol Surv Bull* 311:1–53
- Claire MW, Catling DC, Zahnle KJ (2006) Biogeochemical modelling of the rise in atmospheric oxygen. *Geobiology* 4(4):239–269
- Cloud P (1972) Working model of primitive Earth. *Am J Sci* 272(6):537–548
- Cloud P (1973) Paleocological significance of the banded iron-formation. *Econ Geol* 68:1135–1143
- Crowe SA, Dossing LN, Beukes NJ, Bau M, Kruger SJ, Frei R, Canfield DE (2013) Atmospheric oxygenation three billion years ago. *Nature* 501:535–538
- Embleton BJJ, Williams GE (1986) Low palaeolatitude of deposition for late Precambrian periglacial varvites in South Australia: implications for palaeoclimatology. *Earth Planet Sci Lett* 79:419–430
- Evans DA, Beukes NJ, Kirschvink JL (1997) Low-latitude glaciation in the Palaeoproterozoic era. *Nature* 386:262–266
- Farquhar J, Bao H, Thieme M (2000) Atmospheric influence of Earth's earliest sulfur cycle. *Science* 289:756–758
- Farquhar J, Peters M, Johnston DT, Strauss H, Masterson A, Wiechert U, Kaufman AJ (2007) Isotopic evidence for Mesoarchean anoxia and changing atmospheric sulphur chemistry. *Nature* 449:706–709
- Frei R, Gaucher C, Paulton SW, Canfield DE (2009) Fluctuations in precambrian atmospheric oxygenation recorded by chromium isotopes. *Nature* 461:250–254
- Gaillard F, Scaillet B, Arndt NT (2011) Atmospheric oxygenation caused by a change in volcanic degassing pressure. *Nature* 478:229–232
- Gnos E, Armbruster T, Villa IM (2003) Norrishite, $K(Mn_2^{3+}Li)Si_4O_{10}(O)_2$, an oxymineral associated with sugilite from the Wessels Mine, South Africa: Crystal chemistry and $^{40}Ar-^{39}Ar$ dating. *Am Mineral* 88:189–194
- Goldblatt C, Lenton TM, Watson AJ (2006) Bistability of atmospheric oxygen and the great oxidation. *Nature* 443:683–686
- Goto KT, Sekine Y, Suzuki K, Tajika E, Senda R, Nozaki T, Tada R, Goto K, Yamamoto S, Maruoka T, Ohkouchi N, Ogawa NO (2013) Redox conditions in the atmosphere and shallow-marine environments during the first Huronian deglaciation: insights from Os isotopes and redox-sensitive elements. *Earth Planet Sci Lett* 376:145–154
- Harada M, Tajika E, Sekine Y (2015) Transition to an oxygen-rich atmosphere with an extensive overshoot triggered by the Paleoproterozoic snowball Earth. *Earth Planet Sci Lett* 419:178–186
- Hayes JM (1983) Geochemical evidence bearing on the origin of aerobic biosynthesis, a speculative hypothesis. In: Schopf JW (ed) Earth's earliest biosphere: its origin and evolution. Princeton University Press, Princeton, pp 291–301
- Hilburn IA, Kirschvink JL, Tajika E, Tada R, Hamano Y, Yamamoto S (2005) A negative fold test on the Lorrain Formation of the Huronian Supergroup: uncertainty on the paleolatitude of the Paleoproterozoic Gowganda glaciation and implications for the great oxygenation event. *Earth Planet Sci Lett* 232:315–332
- Hoffman PF, Schrag DP (2002) The snowball Earth hypothesis: testing the limits of global change. *Terra Nova* 14:129–155
- Hoffman PF, Kaufman AJ, Halverson GP, Schrag DPA (1998) Neoproterozoic snowball. *Earth Sci* 281:1342–1346

- Holland HD (1984) The chemical evolution of the atmosphere and oceans. Princeton University Press, Princeton
- Holland HD (2002) Volcanic gases, black smokers, and the great oxidation event. *Geochim Cosmochim Acta* 66:3811–3826
- Isley AE, Abbott DH (1999) Plume-related mafic volcanism and the deposition of banded iron formation. *J Geophys Res-Solid Earth* 104(B7):15461–15477
- Johnson JE, Webb SM, Thomasc K, Onoc S, Kirschvink JL, Fischera WW (2013) Manganese-oxidizing photosynthesis before the rise of cyanobacteria. *Proc Natl Acad Sci* 110:11238–11243
- Karhu J, Holland H (1996) Carbon isotopes and the rise of atmospheric oxygen. *Geology* 24:867–870
- Kasting JF (1993) Earth's early atmosphere. *Science* 259:920–926
- Kasting JF (2005) Methane and climate during the Precambrian era. *Precambrian Res* 137:119–129
- Kasting JF (2013) What caused the rise of atmospheric O₂? *Chem Geol* 362:13–25
- Kasting JF, Egglar DH, Raeburn SP (1993) Mantle redox evolution and the oxidation state of the Archean atmosphere. *J Geol* 101:245–257
- Kasting JF, Pavlov AA, Siefert JL (2001) A coupled ecosystem-climate model for predicting the methane concentration in the Archean atmosphere. *Orig Life Evol Biosph* 31:271–285
- Kirschvink JL (1992) Late Proterozoic low-latitude global glaciation: the snowball earth. In: Schopf JW, Klein C (eds) *The proterozoic biosphere*. Cambridge University Press, Cambridge, UK, pp 51–52
- Kirschvink JL, Kopp RE (2008) Palaeoproterozoic ice houses and the evolution of oxygen-mediating enzymes: the case for a late origin of photosystem II. *Philos Trans R Soc B-Biol Sci* 363(1504):2755–2765
- Kirschvink JL, Gaidos EJ, Bertani LE, Beukes NJ, Gutzmer J, Maepa LN, Steinberger RE (2000) Paleoproterozoic snowball Earth: extreme climatic and geochemical global change and its biological consequences. *Proc Natl Acad Sci* 97:1400–1405
- Kirschvink JL, Gaidos EJ, Bertani LE, Beukes NJ, Gutzmer J, Maepa LN, Steinberger RE (2002) Paleoproterozoic snowball Earth: extreme climatic and geochemical global change and its biological consequences. *Proc Natl Acad Sci* 97:1400–1405
- Klemm D (2000) The formation of Palaeoproterozoic banded iron formations and their associated Fe and Mn deposits, with reference to the Griqualand West deposits, South Africa. *J Afr Earth Sci* 30:1–24
- Konhauser KO, Hamade T, Raiswell R, Ferris FG, Southam G, Canfield DE (2002) Could bacteria have formed the Precambrian banded iron formations? *Geology* 30(12):1079–1082
- Kopp RE, Kirschvink JL, Hilburn IA, Nash CZ (2005) The Paleoproterozoic snowball Earth: a climate disaster triggered by the evolution of oxygenic photosynthesis. *Proc Natl Acad Sci* 102:1131–11136
- Kump LR, Barley ME (2007) Increased subaerial volcanism and the rise of atmospheric oxygen 2.5 billion years ago. *Nature* 448(7157):1033–1036
- Kurzweil F, Wille M, Gantert N, Beukes NJ, Schoenberg R (2016) Manganese oxide shuttling in pre-GOE oceans – evidence from molybdenum and iron isotopes. *Earth Planet Sci Lett* 452:69–78
- Lyons TW, Reinhard CT, Planavsky NJ (2014) The rise of oxygen in Earth's early ocean and atmosphere. *Nature* 506:307–315
- Martin AP, Condon DJ, Prave AR, Lepland A (2013) A review of temporal constraints for the Palaeoproterozoic large, positive carbonate carbon isotope excursion (the Lomagundi–Jatuli event). *Earth Sci Rev* 127:242–261
- North GR, Cahalan RF, Coakley JA (1981) Energy balance climate models. *Rev Geophys Space Phys* 19:91–121
- Ozaki K, Tajika E, Hong PK, Nakagawa Y, Reinhard CT (2017) Effects of primitive photosynthesis on Earth's early climate system. *Nat Geosci*. <https://doi.org/10.1038/s41561-017-0031-2>
- Park JK (1997) Paleomagnetic evidence for low-latitude glaciation during deposition of the Neoproterozoic Rapitan Group, Mackenzie Mountains, N.W.T., Canada. *Can J Earth Sci* 34(1):34–49

- Pavlov AA, Kasting JF (2002) Mass-independent fractionation of sulfur isotopes in Archean sediments: strong evidence for an anoxic Archean atmosphere. *Astrobiology* 2:27–41
- Pavlov AA, Kasting JF, Brown LL (2000) Greenhouse warming by CH₄ in the atmosphere of early Earth. *J Geophys Res* 105:11981–11990
- Pavlov AA, Hurtgen MT, Kasting JF, Arthur MA (2003) Methane-rich Proterozoic atmosphere? *Geology* 31:87–90
- Planavsky NJ, Asael D, Hofmann A, Reinhard CT, Lalonde SV, Knudsen A, Wang X, Ossa F, Pecoits E, Smith AJB, Beukes NJ, Bekker A, Johnson TM, Konhauser KO, Lyons TW, Rouxel OJ (2014) Evidence for oxygenic photosynthesis half a billion years before the great oxidation event. *Nat Geosci* 7:283–286
- Rasmussen B, Buick R (1999) Redox state of the Archean atmosphere: evidence from detrital heavy minerals in ca. 3250–2750 Ma sandstones from the Pilbara Craton, Australia. *Geology* 27:115–118
- Rasmussen B, Bekker A, Fletcher IR (2013) Correlation of Paleoproterozoic glaciations based on U-Pb zircon ages for tuff beds in the Transvaar and Huronian Supergroups. *Earth Planet Sci Lett* 382:173–180
- Rye R, Holland HD (1998) Paleosols and the evolution of atmospheric oxygen: a critical review. *Am J Sci* 298:621–672
- Schroder S, Bekker A, Beukes NJ, Strauss H, van Niekerk HS (2008) Rise in seawater sulphate concentration associated with the Paleoproterozoic positive carbon isotope excursion: evidence from sulphate evaporates in the ~2.2–2.1 Gyr shallow-marine Lucknow Formation, South Africa. *Terra Nova* 20:108–117
- Scott CT, Bekker A, Reinhard CT, Schnetger B, Krapez B, Rumble D III, Lyons TW (2011) Late Archean euxinic conditions before the rise of atmospheric oxygen. *Geology* 39(2):119–122
- Sekine Y, Tajika E, Tada R, Hirai T, Goto KT, Kuwatani T, Goto K, Yamamoto S, Tachibana S, Isozaki Y, Kirschvink JL (2011) Manganese enrichment in the Gowganda Formation of the Huronian Supergroup: a highly oxidizing shallow-marine environment after the last Huronian glaciation. *Earth Planet Sci Lett* 307:201–210
- Shields-Zhou G, Och L (2011) The case for a Neoproterozoic oxygenation event: geochemical evidence and biological consequences. *GSA Today* 21:4–11
- Shields-Zhou G, Porter S, Halverson GP (2016) A new rock-based definition for the Cryogenian period (circa 720 – 635 Ma). *Episodes* 39:3–8
- Summons JR, Jahnke LL, Hope JM, Logan GA (1999) Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* 400:554–557
- Sumner DY, Kirschvink JL, Runnegar BN (1987) Soft-sediment paleo-magnetic fold tests of late Precambrian glaciogenic sediments. *EOS* 68:1251
- Tajika E (2003) Faint young sun and the carbon cycle: implication for the Proterozoic global glaciations, Earth planet. *Sci Lett* 214:443–453
- Zerkle AL, House CH, Cox RP, Canfield DE (2006) Metal limitation of cyanobacterial N₂ fixation and implications for the Precambrian nitrogen cycle. *Geobiology* 4(4):285–297

Chapter 18

End-Paleozoic Mass Extinction: Hierarchy of Causes and a New Cosmoclimatological Perspective for the Largest Crisis



Yukio Isozaki

Abstract The largest mass extinction in the Phanerozoic occurred at the boundary between the Paleozoic and Mesozoic eras (about 252 million years ago). The end-Paleozoic extinction that determined the fate of modern animals including human beings occurred in two steps: first around the Middle-Late Permian boundary (G-LB) and then at the Permian-Triassic boundary (P-TB). Biological and non-biological aspects unique to these two distinct events include changes in biodiversity, isotope ratios (C, Sr, etc.) of seawater, sea level, ocean redox state, episodic volcanism, and geomagnetism. This article reviews possible causes proposed for the double-stepped extinction in regard to the current status of mass extinction studies. Causes of extinction can be grouped into four categories in hierarchy, from small to large scale, i.e., Category 1, direct kill mechanism; Category 2, global environmental change; Category 3, trigger on the planet's surface; and Category 4, ultimate cause. As the G-LB and end-Ordovician extinctions share multiple similar episodes including the appearance of global cooling (Category 2), the same cause and processes were likely responsible for the biodiversity drop. In addition to the most prevalent scenario of mantle plume-generated large igneous provinces (LIPs) (Category 3) for the end-Permian extinction, an emerging perspective of cosmo-climatology is introduced with respect to astrobiology. Galactic cosmic radiation (GCR) and solar/terrestrial responses in magnetism (Category 4) could have had a profound impact on the Earth's climate, in particular on extensive cloud coverage (irradiance shutdown). The starburst events detected in the Milky Way Galaxy apparently coincide in timing with the cooling-associated major extinctions of the Phanerozoic and also with the Proterozoic snowball Earth episodes. As an ultimate cause (Category 4) for major extinction, the episodic increase in GCR-dust flux from the source (dark clouds derived from starburst) against the geomagnetic shield likely determined the major climate changes, particularly global cooling in the past. The study of mass extinctions on Earth is entering a new stage with a new astrobiological perspective.

Y. Isozaki (✉)

Department of Earth Science and Astronomy, The University of Tokyo, Tokyo, Japan
e-mail: isozaki@ea.c.u-tokyo.ac.jp

Keywords Mass extinction · Extinction cause · Large igneous province · Galactic cosmic radiation · Global cooling · Starburst

18.1 Introduction

In the Phanerozoic history of animal evolution for nearly 540 million years, large drops in biodiversity intermittently occurred in multiple times; these are recognized as mass extinction events (Fig. 18.1). Mass extinction is defined as an unusual episode of prominent biodiversity loss among many taxonomic lineages on a global scale and in a geologically short time. Geologic records demonstrate that past mass extinctions apparently occurred immediately before a rapid diversification of new taxa, suggesting paradoxically that mass extinction worked as an accelerator for evolution (Eldredge and Gould 1972). As to the causes of mass extinction, various explanations have been proposed, e.g., a bolide impact for the dinosaur/ammonite-killing Mesozoic-Cenozoic (Cretaceous-Paleogene) boundary (K-PgB) event (ca. 66 million years ago (Ma); Alvarez et al. 1980) and extremely large-scale volcanism for the largest mass extinction in the Phanerozoic at the Paleozoic-Mesozoic (Permian-Triassic) boundary (P-TB) (ca. 252 Ma; e.g., Renne and Basu 1991; Campbell et al. 1992). These two leading hypotheses predominate; nonetheless, numerous unknowns still remain in explaining the causal mechanisms of extinction for a great variety of animals and also the possible link between extinction and the subsequent diversification; thus discussion is ongoing. Except for the debate on ultimate cause(s), a consensus among researchers has been reached that mass extinctions in the past were not of biological origin but were driven by episodic background environmental changes on a global scale from non-biological cause(s) (e.g., Erwin 2006).

During the last two decades, the pattern of major mass extinctions has been updated on the basis of renewed paleontological archives, even for the Big 5, the biggest five major extinctions of the Phanerozoic: end-Ordovician (ca. 444 Ma), Late Devonian (ca. 372 Ma), end-Permian (ca. 252 Ma), end-Triassic (ca. 201 Ma), and end-Cretaceous (ca. 66 Ma) events (Sepkoski 1996; Fig. 18.1a). For example, the end-Triassic extinction is now evaluated as less significant (Lucas and Tanner 2018), whereas the end-Middle Permian event became recognized as more significant than had been previously viewed (Jin et al. 1994; Stanley and Yang 1994; Stanley 2016). Indeed, more than ten mass extinction episodes probably occurred during the Phanerozoic (Bambach 2006). Among these, the P-TB event stands out for the greatest magnitude of biodiversity loss (Arloy et al. 2008; Fig. 18.1b).

This article reviews the current status of mass extinction studies, focusing particularly on the largest of these events, which occurred at the P-TB, and reviews previously proposed explanations. Through this review/summary of remaining issues for future research, a new categorization of causes for mass extinction is proposed. In addition to the conventional explanations, a new perspective of cosmo-climatology is introduced, in order to promote a possible synergic advance between mass extinction studies and astrobiology. Mass extinction is a phenomenon that takes place on the Earth's surface; nonetheless the ultimate cause(s) of the past examples likely came from the Earth's interior and/or its surroundings.

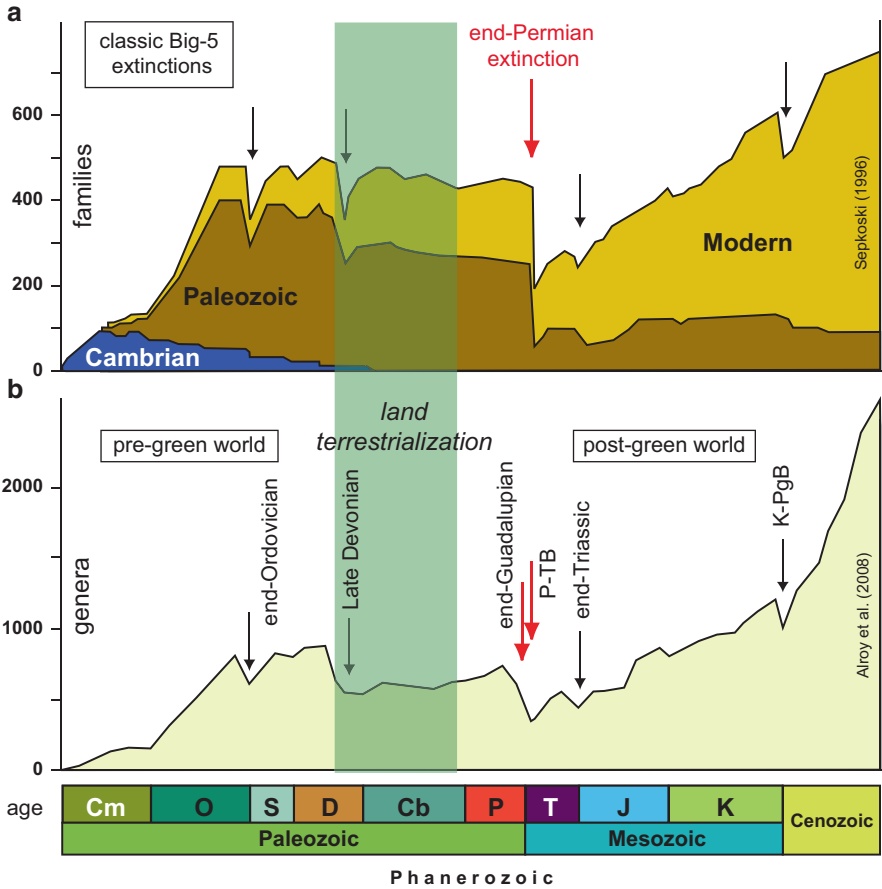


Fig. 18.1 Biodiversity change of marine invertebrates during the Phanerozoic, with emphasis of multiple major mass extinctions. **(a)** The classic compilation by Sepkoski (1996) showing the well-known Big-5 extinction events (shown by arrows). Subdivision of “Cambrian,” “Paleozoic,” and “Modern” represents faunas which dominated in the Cambrian, in the Paleozoic, and in the Mesozoic-Cenozoic, respectively. At the bottom, Cambrian, Ordovician, Silurian, Devonian, Carboniferous, Permian, Triassic, Jurassic, and Cretaceous periods are abbreviated as Cm, O, S, D, Cb, P, T, J, and K, respectively. **(b)** A revised compilation based by Alroy et al. (2008). Note that the end-Paleozoic extinction was the greatest in magnitude among all and that this event was in fact twofold; i.e., the first episode occurred at the Guadalupian-Lopingian boundary (G-LB) and the second at the Permian-Triassic boundary (P-TB) (red arrows in **b**) (Fig. 18.2a, b). The G-LB extinction marked the first major drop in biodiversity of the long-persisted Carboniferous-Permian fauna (refer to Fig. 18.2b). The light green domain shows the interval of mid-Paleozoic terrestrialization by land plants, which divides the Phanerozoic into the “pre-green” world and “post-green” one with contrasting atmospheric composition (refer to Fig. 18.6)

18.2 The Greatest Mass Extinction in History: End-Paleozoic Crisis

18.2.1 Fossil and Stratigraphic Records

18.2.1.1 Victims and Diversity Drop

Fossil records suggest that a large variety of Paleozoic fossil lineages were terminated in a relatively short time interval at the end of the Permian, ca. 252 Ma (Figs. 18.1 and 18.2a). The severe damage selectively hit shallow marine biota, in

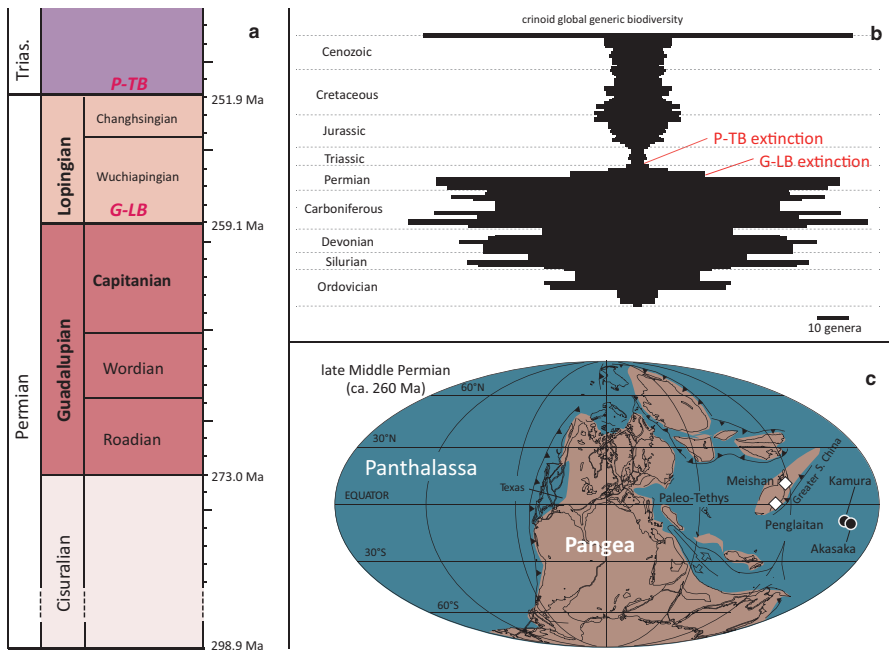


Fig. 18.2 Stratigraphic subdivision of the Permian period (a After subcommission of Permian stratigraphy in Lucas and Shen 2018), the diversity change of crinoid (echinoderm) in the Phanerozoic (b Modified from Baumiller and Messing 2007), and reconstructed Permian paleogeographic map (c Modified from Scotese 2008). Note that the G-LB and P-TB extinction-related boundaries are separated from each other for nearly 7 myr (a) and this separation requires two independent causes and processes of extinction. Diversity change of crinoids throughout the Phanerozoic (b) represents one of the typical double-phased extinction patterns of shallow marine sessile invertebrates of the Permian, i.e., the first major decline at the G-LB after the long-term high diversity and, the second, terminal drop at the P-TB. Detailed information on the extinction-related G-LB and P-TB was obtained from fossiliferous shallow marine strata at the GSSPs (Global Stratotype Section and Points) at Penglaitan for the G-LB and at Meishan for the P-TB in South China and also from accreted mid-oceanic paleo-atoll limestones at Kamura and Akasaka in Japan (c). Significant stratigraphical data sets for the Guadalupian (Middle Permian) were reported from Western Texas with the stratotype sections of the substages (Roadian, Wordian, and Capitanian)

particular sessile benthos, such as rugose corals (cnidaria), fusulines (foraminifers), bryozoans, brachiopods, and echinoderms (e.g., Sepkoski 1996; Alroy et al. 2008; Figs. 18.2b and 18.3a). A remarkable “reef gap” in the geologic history occurred in the aftermath of the P-TB extinction (Kiessling 2001). In contrast, Permian

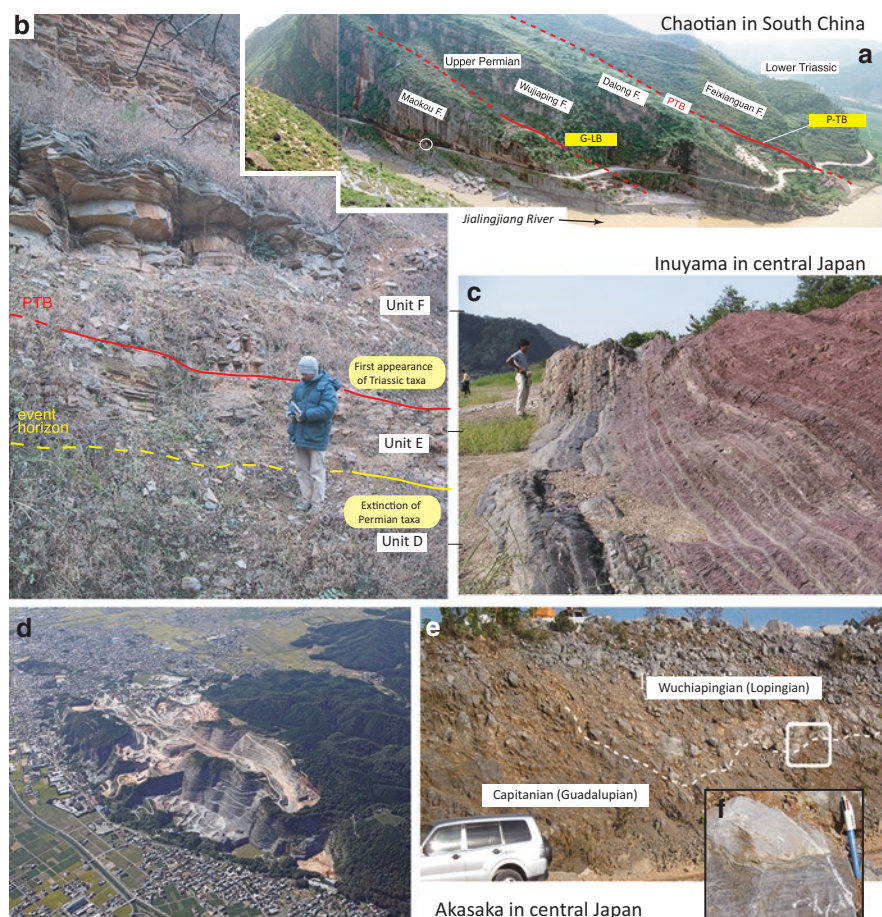


Fig. 18.3 Field views of the extinction-relevant stratigraphic records of the P-TB and G-LB in different lithofacies. (a, b) The G-LB and P-TB intervals of continental shelf-slope facies at Chaotian in Sichuan, South China (Isozaki et al. 2007). The P-TB interval demonstrates two significant horizons, i.e., the extinction horizon of preexisting fauna and the first appearance datum (FAD) of new fauna (b). (c) Post-P-TB interval of mid-oceanic deep sea at Inuyama in central Japan (Isozaki 1997), showing a change in ocean redox, i.e., from anoxic (black pyritic) to oxic (red hematitic) cherts. This outcrop of the Anisian (early Middle Triassic) age demonstrates a recovery interval after the unique P-TB superanoxia episode in the lost superocean Panthalassa (Isozaki 1997). (d–f) The G-LB interval of mid-oceanic atoll facies at Akasaka quarry in central Japan (Kofukuda et al. 2014). Note a sharp hiatus (erosion surface) between the Capitanian (dark gray) and overlying Wuchiapingian (white) limestone (e, f), which recorded a sharp sea-level drop in the low-latitude mid-Panthalassa superocean

locomotive animals and swimmers with high metabolic/respiratory system suffered less (Knoll et al. 1996). This suggests that capability of relocating habitat at the beginning of drastic environmental change, particularly under harsh conditions, divided survivors from victims.

Although estimate of 96% loss on species level (the possible maximum value in statistical prediction; Raup and Sepkoski 1982) has been overemphasized in popular discussions, a realistic and reasonable estimate of the extinction rate is probably about 81% (Stanley 2016). Also on land, severe biodiversity loss occurred in tetrapods and plants (Lucas 2009), and the terrestrial extinction across the P-TB occurred slightly earlier than that of marine biota (Looy et al. 2001).

After the P-TB, most Paleozoic-type biota could never recover the preexisting diversity (Figs. 18.1a and 18.2b), whereas new Modern-type (Mesozoic-/Cenozoic-type) biota including mammals took over the niche and have diversified till present. The P-TB thus marked the biggest singularity point in the Phanerozoic animal evolution.

18.2.1.2 Two Steps

The putative biggest extinction of the Phanerozoic occurred in fact in two distinct steps that are separated from each other by seven million years, i.e., first near the Middle-Late Permian (Guadalupian-Lopingian) boundary (G-LB; ca. 259 Ma) and second at the P-TB *per se* (ca. 252 Ma) (Stanley and Yang 1994; Jin et al. 1994; Figs. 18.1b, 18.2a, b, and 18.3a, b). The second extinction at the P-TB *sensu stricto* is regarded as the severest biodiversity loss in the Phanerozoic, and debates still continue on how many extinction pulses occurred.

Although the G-LB extinction marks the first significant decline in Carboniferous-Permian shallow marine biota (Fig. 18.2b), this episode was overlooked until the 1990s, owing to the great shadow of the P-TB event. Nearly 62% of shallow marine animals were terminated (Stanley 2016), in particular rugose corals, large-tested fusulines, bryozoans, crinoids, brachiopods, ammonoids, and other tropically adapted fauna in low latitudes (e.g., Stanley and Yang 1994; Wang and Sugiyama 2000; Isozaki and Aljinovic 2009). In addition, during the late Middle Permian, mid-latitude brachiopod fauna started to migrate into lower-latitude areas (Shen and Shi 2002). However, the extinction was likely not an acute event, but rather a prolonged episode that occurred throughout the Capitanian (= late Guadalupian; Fig. 18.2a) (Clapham et al. 2009). Although not fully documented yet, land biota also suffered a major decline in diversity (Lucas 2009; Rubidge et al. 2013). Although the Late Permian witnessed a moderate but not full recovery of biodiversity after the G-LB event, the G-LB and P-TB events are clearly distinguished.

18.2.1.3 Timing

Owing to the recent development in high-precision U-Pb dating for individual zircons from intercalated tuff beds, extinction-related horizons around the G-LB and P-TB were constrained in age much better than before. The P-TB extinction likely occurred between 251.941 ± 0.037 and 251.880 ± 0.031 Ma, as confirmed at the world's standard section (Global Stratotype Section and Point, GSSP) at Meishan in South China (Burgess et al. 2014; Fig. 18.2a, c). Regardless of how many extinction peaks existed, the P-TB extinction likely occurred within a short time interval of nearly 60,000–48,000 kyr. In contrast, the G-LB event was a prolonged process that spanned the Capitanian (Clapham et al. 2009). The age of the biostratigraphically defined G-LB is now assigned to be 259.1 ± 0.5 Ma (Shen et al. 2013).

It is noteworthy that the two extinctions are independent and that they were separated from each other by ca. 7 myr (Fig. 18.2a), because these require two independent triggers and two relevant environmental changes of global scale. In the following sections, major records of global phenomena that coevally occurred around the G-LB and P-TB timings are summarized in two parts, i.e., biogeochemical signatures and non-biological ones.

18.2.2 Biogeochemical Records

During the last 20 years, the practical utilization of various new geochemical proxies for analyses of paleoenvironments advanced dramatically. In addition to conventional analyses of major and trace element composition, advances in the utility of REEs (rare-earth elements), various isotope composition metrics, and biomarkers are particularly significant. For the end-Paleozoic extinctions, numerous data have been likewise published during the last 20 years.

18.2.2.1 Perturbation in C Cycle

Carbon, sulfur, and nitrogen isotope ratios in ancient sedimentary rocks, in particular in carbonates, recorded biogeochemical cycles of these bio-essential elements in the surface environments, which reflect the status and long-term changes of the Earth's biosphere. As these geochemical proxies provide constraints on ancient bio-productivity and the burial rate of organic material in local water masses and also on the global ocean, the measurements of these proxies became a conventional approach in studies on some extinction-related unique intervals in the Phanerozoic (Kump and Arthur 1999; Saltzman and Thomas 2012; Paytan and Gray 2012).

Volatile fluctuations in carbon isotope ratio of seawater generally occurred during particular intervals around major extinction timings in the Phanerozoic. As to the P-TB event, a sharp negative shift in the carbon isotope ratio of carbonates was

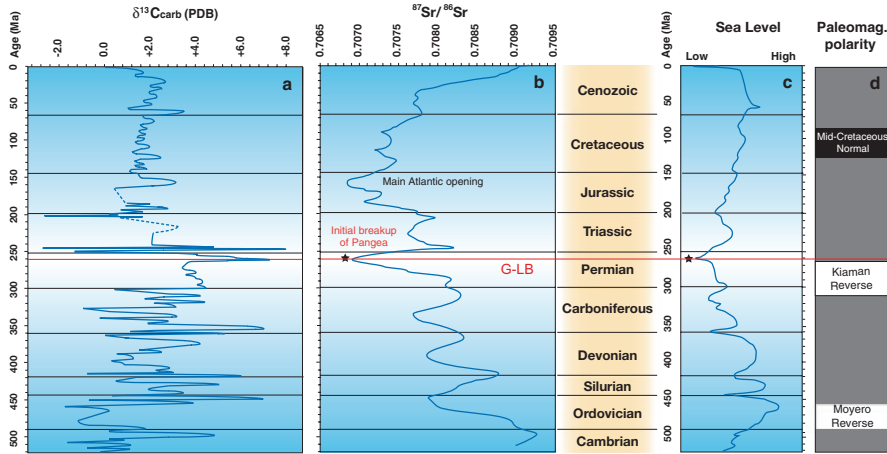


Fig. 18.4 Compilation of Phanerozoic secular changes in inorganic carbon isotope ratio ($\delta^{13}\text{C}_{\text{carb}}$) and strontium isotope ratio ($^{87}\text{Sr}/^{86}\text{Sr}$) of seawater (**a**, **b**, respectively), sea level (**c**), and geomagnetic polarity (**d**) (Modified from Isozaki 2009a, b). Among these, only $\delta^{13}\text{C}_{\text{carb}}$ of seawater was related to biological activities, whereas the rest reflected non-biological phenomena. **(a)** The onset of volatile changes of C isotope across the Paleozoic-Mesozoic transition occurred not at the P-TB but in the Capitanian (Middle Permian) immediately before the G-LB, which is marked by a positive excursion called the Kamura event (Isozaki et al. 2007). **(b)** The minimum value of Paleozoic seawater Sr-isotope ratio occurred in the Capitanian, not around the P-TB (Korte et al. 2005; Kani et al. 2013). **(c)** The lowest sea level of world oceans occurred around the G-LB not at the P-TB (Haq and Schutter 2008). **(d)** Long-term geomagnetic stability (Kiaman Reverse Polarity Superchron) collapsed immediately before the G-LB (Steiner 2006; Isozaki 2009a, b; Kirschvink et al. 2015). Note that non-biological global phenomena recorded a unique singularity point not at the P-TB timing but at the G-LB in the Permian. Biological phenomena (fluctuation in C-isotope ratio of seawater and extinction) likely followed a global-scale major environmental change recorded by these. Almost the same observation is made for the end-Ordovician extinction—involving event with similar signatures in C isotope, Sr isotope, sea level, and geomagnetic polarity stability/reversal (Isozaki and Servais 2018)

detected at the P-TB horizon almost in all analyzed sections in the world (Holser and Magaritz 1987; Musashi et al. 2001; Korte et al. 2005; Shen et al. 2013; Figs. 18.2c and 18.4a). This indicates an abrupt and large flux of isotopically light carbon into world oceans, likely suggesting a collapse of primary productivity on a global scale, in other words, a significant malfunction of the global food web. On the other hand, a unique positive excursion of the carbon isotope ratio of carbonates of up to $+7\text{‰}$ (the Kamura event) was detected in the Capitanian, which was followed by a sharp negative drop of ca. 5‰ immediately below the G-LB horizon (red horizontal line in Fig. 18.4; Wang et al. 2004; Isozaki et al. 2007, 2011). These unique signals at and around the major extinctions suggest a significant change in the carbon cycle in the surface ocean.

In addition, isotopic records of organic carbon, sulfur, and nitrogen isotope essentially confirmed the appearance of acute perturbation in the biosphere both

around the G-LB and P-TB (e.g., Kaiho et al. 2001; Cao et al. 2009; Shen et al. 2011; Saitoh et al. 2014), such as the expansion of an oxygen minimum zone in ocean with a big tongue of anoxic water at the shelf edge.

18.2.2.2 Lipid Biomarkers and Other Organic Molecules

Some biologically synthesized large molecules in unmetamorphosed sedimentary rocks are called biomarkers because they are useful in detecting the sources of organic matter. For extinction-related intervals in the geologic past, the utility of various biomarkers has been much emphasized lately (Whiteside and Grice 2016). The P-TB extinction horizons are characterized by some biomarkers such as aromatic molecules (e.g., isorenieratane and maleimides derived from pigments, isorenieratene and bacteriochlorophylls of Chlorobi), which indicate the development of extremely oxygen-depleted conditions even in the euphotic shallow sea (e.g., Grice et al. 2005). In addition, other biomarkers were detected also from the P-TB interval, such as dibenzothiophene, dibenzofuran, and biphenyl derived from plant lignin, that suggest abundant flux of land plant remains into sediments and background terrestrial soil erosion (Xie et al. 2007). The occurrence of some polycyclic aromatic hydrocarbons (PAHs) also supports a high-temperature combustion episode like forest fire at the P-TB (Nabbefeld et al. 2010). These lines of biomarker evidence equally indicate the severe deterioration of marine and terrestrial ecosystems for the specific timings related to the P-TB extinction episodes. As to the G-LB, more work is needed.

18.2.3 *Non-biological Global Phenomena*

Apart from abovementioned phenomena related to biological activities, the following unique non-biotic phenomena, all on global scale, occurred at the end of the Paleozoic, i.e., remarkable sea-level drop, minimum in seawater Sr-isotope ratio, violent volcanism, and end of paleomagnetic superchron (an interval characterized by a uniquely long-term stable geomagnetic polarity) (Fig. 18.4). Judging from their global aspects, all of them probably were related with the major change in environments and biota, which might possibly have changed the course of evolution.

18.2.3.1 Lowest Sea Level

The sea-level changes around the Paleozoic-Mesozoic transition are unique in the Phanerozoic. As clearly observed in the long-term change in global sea level based on the compilation of sequence stratigraphy, the lowest sea level of the Phanerozoic was recorded across the G-LB (Fig. 18.4c; Haq and Schutter 2008). The magnitude

of the sea-level drop is estimated as more than 100 m, and the lowest level was much lower than that in the Gondwana glaciation in the Late Carboniferous and Early Permian. In contrast, the P-TB timing corresponds to the middle of a long-term sea-level rise after that. As observed in many sections in the world, Middle Permian strata are indeed unconformably covered directly by Lower Triassic without Upper Permian beds. The almost total absence of continuous sedimentary records across the two significant extinction horizons makes the GSSP for the G-LB at Penglaitan and that for the P-TB at Meishan both in South China (Fig. 18.2c) exceptionally valuable. The unconformity was found also in a mid-oceanic paleo-atoll limestone that was deposited on top of the seamount in the ancient mid-ocean (Akasaka in Figs. 18.2c and 18.3C–F; Kofukuda et al. 2014). The finding suggests that the sea-level drop at the end-Middle Permian was caused not by a local tectonic uplift of seamount (extremely rare in mid-ocean) but by the global sea-level change (Fig. 18.4c).

These observations suggest the transfer of a large quantity of seawater from the oceans onto land in the form of ice; otherwise such a large-scale drop in global sea level cannot occur, thus suggesting the cold climate in the G-LB interval. Nonetheless, extensive glacial deposits, e.g., tillite/diamictite, were not detected in the G-LB interval, except for a minor amount of high-altitude alpine-type mountain glaciers (Fielding et al. 2008). This apparent disagreement needs to be examined. Such a global sea-level drop was unique to the G-LB but not to the P-TB (Fig. 18.4c); this suggests different nature of environmental changes and also of cause between the two major extinctions.

18.2.3.2 Capitanian Sr Minimum

The isotopic ratio of strontium ($^{87}\text{Sr}/^{86}\text{Sr}$) in seawater reflects the balance between two fluxes, i.e., input of relatively light-mass Sr from hydrothermal activity along mid-oceanic ridges and continental flux of eroded continental crust material enriched in radiogenic heavier Sr (McArthur et al. 2012). The seawater Sr-isotope ratio is recorded in carbonates. The long-term change in the Phanerozoic (Veizer et al. 1999; Fig. 18.4b) clearly shows a minimum not only of the Permian but also of the entire Phanerozoic. The minimum value of seawater Sr ratio ($^{87}\text{Sr}/^{86}\text{Sr} = \text{ca. } 0.7068$; Veizer et al. 1999; Korte et al. 2006) occurred in the Capitanian, immediately before the G-LB, whereas no significant change is detected across the P-TB. This signature called the Capitanian minimum (Kani et al. 2013) indicates that the Capitanian seawater was characterized uniquely by the least continental flux with respect to that from mid-oceanic ridge in the Phanerozoic.

In general, the total activity of mid-oceanic ridges is and has been stable as long as plate tectonics has operated throughout the Phanerozoic. Thus the relatively low and the lowest values of Sr isotopes during the Capitanian suggest a significant shutdown of continental flux into the world ocean. There are three to cause short-term suppression of continental flux on global scale, i.e., (1) an extremely large rise in sea level to drown continents extensively, (2) decline in riverine flux by predomi-

nance of arid climate over continents (Korte et al. 2006), and (3) regional ice covering over continental crust (Kani et al. 2013). Judging from the extremely low sea level across the G-LB, the first option is unlikely.

18.2.3.3 Anoxia/Euxinia

In the early 1990s, the ubiquitous occurrence of black shales from the near-P-TB interval was revealed from shallow marine facies (Wignall and Hallam 1992). These unique strata commonly contain high TOC (total organic carbon) and numerous framboidal pyrite (FeS_2), and these suggest the deposition under oxygen-depleted (anoxic) conditions and the appearance of an unusual environmental condition in continental shelves around supercontinent Pangea. On the other hand, deep-sea bedded cherts with the peculiar black mudstone across the P-TB horizon were found in Japan, which feature similar anoxic signature (Isozaki 1997; Fig. 18.3c). These cherts were deposited primarily in deep sea below the carbonate compensation depth (CCD; ca. 3000 m deep) in mid-oceanic setting and were secondarily accreted to the subduction-related continental margin of Japan in the Middle Jurassic (Matsuda and Isozaki 1991): they recorded information on the extinction-related, global environmental changes in the lost superocean Panthalassa (Isozaki 2014; Fig. 18.2c). The P-TB anoxic black mudstone/chert is stratigraphically sandwiched by under- and overlying red cherts that represent a contrasting oxic setting, suggesting the appearance of unusual conditions of long-term anoxia (superanoxia) across the P-TB in the superocean.

Recent updates in microfossil biostratigraphy and in redox-sensitive elemental analyses revised the total duration and nature of the event. Nonetheless the P-TB superanoxic episode represents one of the most significant changes in global ocean redox state during the Phanerozoic, implying a major impact to the biotic crisis across the P-TB, because most animals in sea and on land depend on oxygen respiration. In contrast to the previous explanation of total stratification of the ocean, a significant expansion of the oxygen minimum zone over the entire superocean was proposed (Algeo et al. 2011). In coeval shallow marine continental shelves, strongly reduced conditions (anoxia and even euxinia with H_2S) may have appeared by forming a big tongue of anoxic water at the shelf edge.

Regarding the G-LB interval, another anoxic signature was detected in the Capitanian marine sequence of shelf/slope facies in South China (Saitoh et al. 2013; Shi et al. 2016); nonetheless the total duration of shelf anoxia was much shorter than the P-TB superanoxia.

18.2.3.4 Flood Basalt-Large Igneous Province (LIP)

One of the most spectacular non-biological episodes across the P-TB was the eruption of ca. 252 Ma Siberian Traps, which is composed of an extremely thick pile of Ti-rich, alkaline, mafic (rich in Fe^{2+} , Fe^{3+} , and Mg^{2+}) lava flows (up to 6500 m thick)

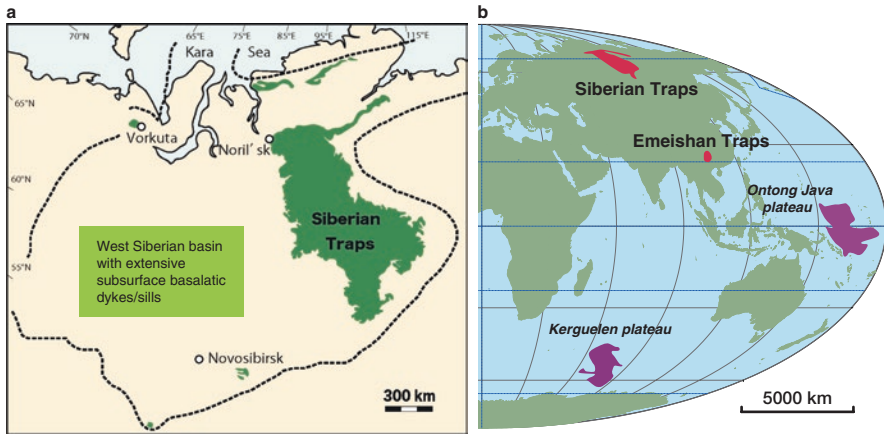


Fig. 18.5 The Siberian Traps formed around the P-TB timing (simplified from Saunders et al. 2005; Ernst 2014). (a) Surface occurrence of effusive parts of the Siberian Traps (basaltic lava and pyroclastics in green domain) with the presumed extent of subsurface distribution of relevant dykes/sills (area within a broken line). (b) Locations of the two end-Paleozoic LIPs, i.e., the near-P-TB Siberian Traps in Russia and the near-G-LB Emeishan Traps in South China. For comparison, two major oceanic LIPs (Ontong Java and Kerguelen oceanic plateaux) are also shown in violet color. Note that the sizes of the oceanic LIPs are much larger than that of the two continental LIPs; however, no extinction occurred at the eruption of the former

with extensive magmatic intrusions of mafic/ultramafic composition developed underneath (>1.6 million km^2 ; Federenko et al. 2000; Reichow et al. 2009; Svensen et al. 2009; Fig. 18.5a). The eruption likely occurred in a highly limited time less than 1 myr (Burgess et al. 2017). The total volume, geochemical characteristics, and rapid eruption indicate their origin, not in subduction-related volcanic arcs nor mid-oceanic ridges, but in a large igneous province (LIP) induced by a mantle plume (e.g., Saunders et al. 2005). Among the known LIPs in the world, the Siberian Traps is one of the largest in volume (Ernst 2014).

Already in the early 1990s, some pioneer researchers realized the coeval timing between the flood basalt volcanism and the P-TB extinction and suggested a possible cause-effect link (Renne and Basu 1991; Campbell et al. 1992; Kamo et al. 1996; Courtillot 1999; Wignall 2001). This possibility has been much emphasized later due to more precise dating of the Siberian Traps *per se* and also of fossil-bearing P-TB strata (e.g., Shen et al. 2013; Burgess et al. 2017). The associated thick Neoproterozoic-Cambrian oil-bearing strata and evaporites in the same area have been emphasized, because they were likely intruded by the end-Permian sills of unusually high temperatures to have caused the emission of large volume of CO_2 (Svensen et al. 2009). Despite the claimed large scale, however, direct

evidence for volcanism, such as mafic scoria and other volcanoclastics, has not been reported in the rest of the world outside Siberia. Exceptions are the occurrence of coal fly ash and spikes of Hg concentration in the near P-TB beds, which are regarded as products derived directly from the Siberian Traps (Grasby et al. 2011, 2015); nonetheless the putative effects to the biosphere and biodiversity need further evaluation.

Likewise, the Emeishan Traps in South China has been focused (e.g., Chung et al. 1998; Wignall et al. 2009; Bond et al. 2010; Fig. 18.5b), because its eruption timing was close to the end-Paleozoic extinction, in particular to the G-LB event. The size of the Emeishan Traps, however, is significantly small with respect to other major LIPs. It erupted in a short time from 259.6 to 257.6 Ma, mostly after the G-LB (259.1 Ma) (e.g., Shellnutt 2014; Zhong et al. 2014); thus it unlikely caused either the Capitanian cooling or the prolonged Capitanian extinction. Though a Capitanian basalt lava in Yunnan was claimed as a precursor of the Emeishan Traps (Wignall et al. 2009), it is far too small to drive major extinction.

18.2.3.5 Illawarra Reversal

During the Phanerozoic, three distinct intervals of paleomagnetic uniqueness are recognized, i.e., mid-Cretaceous Normal, Carboniferous-Permian Kiaman Reverse, and Ordovician Moyero Reverse superchrons (Isozaki 2009a; Fig. 18.4d). The rest of the Phanerozoic is characterized by frequent geomagnetic polarity changes, whereas these three intervals witnessed a long-term stability in polarity over many million years. Stable geomagnetic polarity for such a long time suggests the stable convection in the outer core of the Earth where geomagnetism is generated by the geodynamo.

Regarding the Permian case, the end of the Kiaman superchron is called the Illawarra Reversal after the locality name of the pioneering paleomagnetical study in East Australia (Irving and Parry 1963). This episode occurred in the Wordian (middle Guadalupian), before the Capitanian extinction and G-LB (Steiner 2006; Isozaki 2009a; Kirschvink et al. 2015; Belica et al. 2017; Fig. 18.4d). This indicates that the geodynamo changed its mode from a highly stable to more frequently fluctuating condition; in other words, convection in the outer core changed from a dynamically stable mode to a relatively unstable one. The heterogeneity in physical conditions of lower mantle, in particular the change in temperature and/or thermal gradient, is needed to change the mode of geodynamo in the outer core. This change in the deep interior of the planet was connected also to the origin of mantle plume; thus an episodic birth/uprising of a large-scale mantle plume possibly linked a major change along the core-mantle boundary to those on the surface of the planet, which will be discussed later.

18.3 Possible Causes and Remaining Issues

18.3.1 *Categories of Causes*

Various possible causes have been proposed to date for the P-TB and G-LB extinctions as well as for other extinction events in the Phanerozoic (Fig. 18.1), e.g., global cooling/warming, bolide impact, LIP volcanism, and anoxia on a global scale. The word “cause,” however, has been used ambiguously for many years among researchers without a clear definition. To minimize confusion in this debate, which has lasted over a century, I propose here to group previously claimed causes of mass extinction into the following three distinct categories in hierarchy from small to large scale, namely, Category 1, direct killing mechanism; Category 2, background environmental change on a global scale; and Category 3, main trigger appearing on the Earth’s surface; and to establish Category 4 for ultimate cause (Table 18.1).

Causes of Category 1 include various kill mechanisms that are responsible for terminating individual animal group dwelling in a particular living habitat of local/regional but not global context, e.g., temperature drop/rise, changes in humidity, water salinity/pH/redox, toxication, etc., which may drive malfunction in metabolism and/or collapse of nerve systems with dehydration, suffocation, hypercapnia, nutrient shortage, metal poisoning, etc. (Table 18.1). Some kill mechanisms are related also to ecological structure of animal communities, such as predation, thus biological in nature, whereas most others are purely non-biological. Judging from the post-extinction recovery processes/results, these causes are relatively short-lived, and damages are restorable. As life and the surrounding environments on Earth are and have been so diverse, multiple kill mechanisms need to operate together in order to drive mass extinction.

Those of Category 2 represent much larger-scale, global changes in surface environment, in particular climate changes, such as global cooling/warming (Table 18.1), which can generate various kill mechanisms of Category 1. Major glaciation is a typical phenomenon that can change regional landscapes of the biosphere, which can lead modification/destruction of ecological structures for preexisting life on the Earth’s surface. As to the Big-5 extinctions (Fig. 18.1), relevant climate changes were of long time range, much longer than the Quaternary Milankovitch-tuned glacial-interglacial cycles. Our planet intermittently experienced major glaciations in the past, not only multi-times in the Phanerozoic but also twice or more in the Proterozoic, i.e., snowball Earth events (Hoffman and Schrag 2002). In addition, a long-term drop in solar irradiance also caused a profound impact on the photosynthesis-based ecosystem on the Earth’s surface. Rapid diversifications of new lineages or macroevolutions, such as the first appearances of eukaryotes and vertebrates, occurred immediately after the prominent cold spikes, probably associated with cryptic extinctions of preexisting taxa. These significant changes in biosphere, i.e., environmental settings, ecology, and biota, are fatal, which belong to causes of Category 2 that are non-biological in nature.

Table 18.1 Four categories of extinction causes

Category	Processes	Victims
1. Kill mechanism	Temperature drop	Most biota
	Temperature rise	Warmwater dwellers
	Oxygen depletion	Animals with high metabolism
	Hypercapnia	Animals with low metabolism
	Metal toxicity	Animals with complex nerve system
	pH drop	Calc. shell-forming organisms
	Aridity rise	Aquatic animals/plants on land
	Dust/aerosol screen	Plants and photosynthetic bacteria
2. Global environmental change	Cooling/glaciation	Mid-latitude/tropical fauna
	Warming	Tropical fauna
	Irradiance drop	All biota
3. Trigger on surface	Bolide impact	All biota
	LIP-mantle plume	All biota
	Megaflux of GCR	All biota
4. Ultimate driver		
Terrestrial agents	Mantle convection	All biota
	Core convection	All biota
Extraterrestrial agents	Starburst	All biota
	Supernovae	
	Active galactic center	
	Dark cloud (nebula)	

Possible causes of mass extinction are classified into four categories: Category 1 to Category 4 from small to large scale. Category 1 includes direct kill mechanisms for each biota rather on local basis, whereas Category 2 comprises global-scale environmental changes that induce various kill mechanisms. Category 3 represents major trigger of global environmental changes, which episodically appear on the planet's surface. Those grouped into Category 4 are *bona fide* ultimate causes that originate not on the Earth's surface but in its interior and/or in the outer space. Most causes previously proposed belong to Categories 1–3, and those corresponding to Category 4 have been rarely explored

Category 3 includes triggers for significant course changes in global climate and conditions of biosphere. Solely non-biological large-scale agents can trigger such irreversible climate changes of global context (Table 18.1). A good example of causes of Category 3 is a large-scale impact of extraterrestrial bolide, which is much discussed for the K-Pg event. In addition, unusually large-scale volcanism at LIP is another major cause of Category 3, as debated for the P-TB event. These large-scale non-biological phenomena, in terms of acute influxes of material and energy, can cause mass extinction by renewing the surface environment in various ways and prepare vacant niches for subsequent bio-diversification.

Similar efforts to classify causes of extinction have already been attempted several times; however, what were claimed previously as “ultimate causes” merely imply large-scale triggers that appeared on the Earth's surface (e.g., Bond and Grasby 2017); therefore, both bolide impact and LIP activity correspond merely to causes of Category 3. These unusual “whole-Earth” events are generally regarded to

have occurred randomly thus accidentally; however, the planet Earth is composed of various components that are seamlessly connected to each other, e.g., core, mantle, crust, and hydro- and atmosphere, and the biosphere is limited to the planet's surface. For generating a sufficient extinction trigger episodically on the Earth's surface (causes of Category 3), the neighboring components cannot stay irrelevant; instead, something deep in the planet's interior or in exterior must be connected and responsible for the change. Causes of Category 4 (Table 18.1) are *bona fide* ultimate drivers of large-scale catastrophes eventually appearing on the Earth's surface.

18.3.2 Proposed Causes for the P-TB and G-LB Events

Regarding the P-TB and G-LB events, various explanations have been hitherto proposed: a drop in seawater temperature, depletion in oxygen in seawater (anoxia, hypercapnia), toxic metal poisoning, drop in seawater pH (acidification), etc. All of these kill mechanisms are grouped in Category 1 that may occur essentially on a local basis (Table 18.1). Nonetheless, for terminating a great variety of biota all over the world, mass extinction requires multiple kill mechanisms that appear simultaneously within a limited time period. Global cooling is one of the classical explanations for the cause of extinction (e.g., Stanley 1988), and global warming becomes recently popular by analogy to the modern world (e.g., Sun et al. 2012; Benton and Newell 2014). These are grouped into Category 2, as they can be related potentially to the claimed causes of Category 1.

In contrast, the explanations of a bolide impact (Becker et al. 2001; Kaiho et al. 2001) and of Siberian/Emeishan LIP-Traps (e.g., Renne and Basu 1991; Campbell et al. 1992; Chung et al. 1998; Svensen et al. 2009; Burgess et al. 2017) obviously belong to Category 3 (Table 18.1). The idea of bolide impact at the P-TB suffered serious criticisms for every line of evidence presented (Farley and Mukhopadhyay 2001; Isozaki 2001; Koeberl et al. 2002; Farley et al. 2005), and there have been no proposals since then.

On the other hand, various relevant processes of Categories 1–3 were examined/confirmed, and also their possible mutual links were modeled in various flow charts centered by the Siberian LIP-Traps at P-TB (e.g., Hallam and Wignall 1997; Bond and Grasby 2017). The Siberian Traps produced a huge amount of high-temperature basaltic lava flows (Federenko et al. 2000; Saunders et al. 2005; Reichow et al. 2009; Fig. 18.5) and probably caused direct damage to the surrounding area in the form of an extensive forest fire. Svensen et al. (2009) proposed that the volcanic gas emitted with huge amount of CO₂, together with baking coal-bearing strata to increase CO₂ of the atmosphere. The accumulation of a high level of atmospheric CO₂ has possibly driven super-greenhouse effect to induce global warming and associated ocean acidification (Retallack 2013). In addition, the sluggish ocean circulation, under a minimized thermal gradient between tropical and polar regions, may have allowed the development of long-term anoxic conditions (superanoxia).

More or less the same scenario was proposed for the G-LB extinction and the Emeishan Traps in South China (Bond et al. 2010).

Candidates for Category 4 have been rarely proposed both for the P-TB and G-LB extinction events. The geological lines of evidence indicate that both the Siberian and Emeishan Traps had been formed due to a mantle plume activity (e.g., Ernst 2014); however, the reason why the LIP formation (cause of Category 3) occurred in end-Permian Siberia and end-Guadalupian southwestern South China has not been fully explained, except for the speculation of the “plume winter scenario” (Isozaki 2009b).

18.3.3 *Remaining Conundrum*

As introduced above, the violent volcanism of the LIP-Traps magmatism is currently regarded as the leading explanation for the cause of the end-Paleozoic double extinction events at the P-TB and G-LB (e.g., Wignall et al. 2009; Bond and Grasby 2017); however, we must realize that the ultimate cause (of Category 4; Table 18.1) has not yet been identified. In the following, the essential issues remaining in the P-TB extinction study are listed.

There are three major points to be checked for explaining the cause-effect relationship between LIP and extinction event, i.e., (1) the extent of volcanic hazards due to LIP formation, (2) nature of global climate change, and (3) volcanogenic CO₂-related greenhouse effect.

First of all, the sizes of the Siberian and Emeishan Traps are not necessarily the largest among all LIPs on the planet when compared with the Ontong Java Plateau in SW Pacific (Fig. 18.5). These plumes did not contribute to supercontinent breakup (of Pangea; Fig. 18.2c) like other major LIPs, neither. A large-scale volcanic eruption of the Siberian Traps (Fig. 18.5) might destroy ecosystem immediately around the volcanic center by spreading heat and fire. Nonetheless, expected direct influences to other lands, particularly to those on the other hemisphere and to the vast superocean (Panthalassa) (Fig. 18.2c), have not been documented and thus need to be examined. There is no obvious material input from the Siberian Traps to Panthalassa except for trace spikes of Hg (Grasby et al. 2015). The total abundance of Hg in the Earth’s crust and mantle, nevertheless, appears too dilute to cause metal poisoning for extinction even in regional context.

Second, proponents of LIPs-extinction link explained volcanogenic climate change and effects to biosphere in two ways; i.e., one is global cooling by assuming dust/aerosol screen (volcanic winter) for the G-LB Emeishan Traps, and the other is global warming by elevated partial pressure of atmospheric CO₂ (volcanic summer) for the P-TB Siberian Traps (e.g., Bond and Grasby 2017). These two interpretations in fact suggest the opposite effects to biospheric environments by similar volcanism. Particularly for the G-LB and P-TB episodes, the claimed LIP activity and consequences need more consistent explanations without ad hoc assumptions.

Though the apparent temporal coincidence between volcanism and extinction can promote much discussion, it has not proved the claimed cause-effect link.

The most critical obstacle is the third one, i.e., the background atmospheric CO₂ during the Permian against the putative volcanogenic greenhouse effect. The latest comparison between the G-LB and the end-Ordovician extinction events (Fig. 18.1) pointed out that these two episodes shared more similarities than differences, e.g., extinction pattern of tropical biota, global cooling with >100 m sea-level drop, the unique signature in Sr and C isotope of seawater, and pattern of paleomagnetic polarity change (Isozaki and Servais 2018; Fig. 18.4). This suggests that the same trigger and processes of global scale have driven global cooling and relevant extinction for both cases. Nevertheless, there are crucial differences between the two; one is the absence of LIPs at the end-Ordovician, and the other is the extremely high atmospheric CO₂ level in the Ordovician (Fig. 18.6).

According to the climate modeling by Berner (2006) and Royer et al. (2014), the atmospheric CO₂ content in the Ordovician is estimated about 2500 ppm, whereas that in the Permian was no more than 300 ppm, nearly one order of magnitude lower than in the Ordovician (Fig. 18.6). This significant difference was led obviously by the rapid terrestrialization during the mid-Paleozoic, i.e., the photosynthetic revolution by land plants and development of forests (Fig. 18.1). The global sea-level drop and global cooling with glaciation of the same order (Fig. 18.4c) cannot be reproduced particularly under the high CO₂ level in the Ordovician. In contrast to the volcanism-driven increase in atmospheric CO₂ level, there is no effective mechanism on Earth for quick drawdown of large quantity of atmospheric CO₂ to suppress greenhouse effect in a geologically short time. This casts doubt on the fundamental assumption that the change in atmospheric CO₂ was the sole main driver of the Paleozoic climate and relevant extinction and implies that we need an alternative explanation.

18.4 Ultimate Cause (Category 4)

18.4.1 *Cosmoclimatological Driver*

“Cosmoclimatology” focusing on the influence of galactic cosmic radiation (GCR) on long-term change in the Earth’s climate (Svensmark and Friis-Christensen 1997; Shaviv and Veizer 2003; Svensmark and Calder 2007) is an emerging and promising paradigm for explaining extinction cause of Category 4 (Table 18.1). At the end of this review, we explore this new perspective that is totally different from preexisting viewpoints/scenarios.

Besides the commonly discussed bolide impact scenario, there was a classic extraterrestrial explanation proposed for extinction: i.e., a possible link between a supernovae explosion in the neighborhood of our Solar System and radiation-induced extinction (Schindewolf 1955); however, with respect to the knowledge in

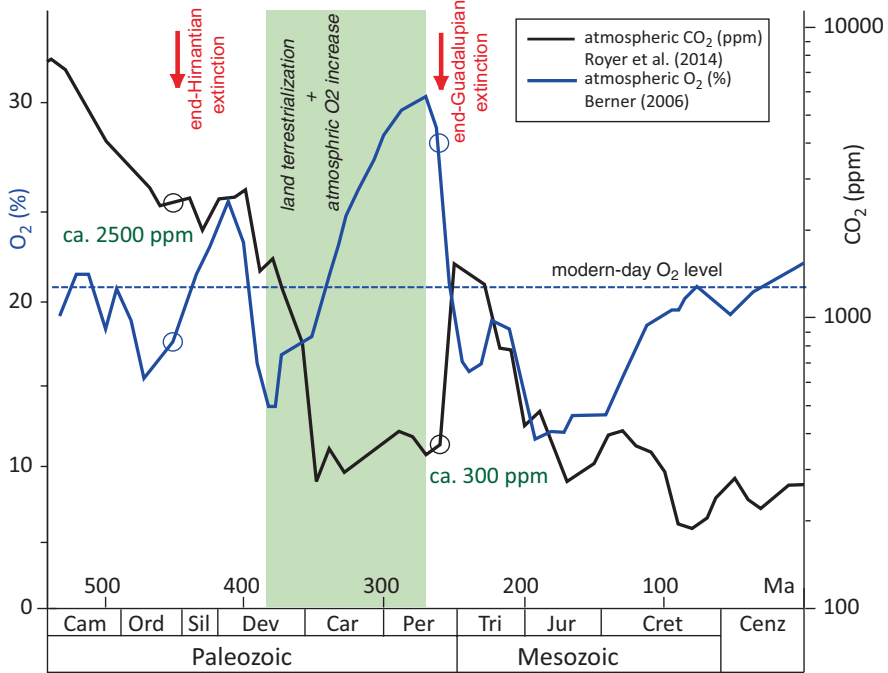


Fig. 18.6 The long-term change in atmospheric CO_2 and O_2 during the Phanerozoic (modified from Isozaki and Servais 2018, compiled from Berner 2006; Royer et al. 2014). Note the remarkable difference in CO_2 level before and after the mid-Paleozoic terrestrialization by land plants (Fig. 18.1), which divided the Phanerozoic into the CO_2 -enriched pre-green world and O_2 -enriched post-green one. The end-Ordovician and end-Guadalupian (G-LB) extinctions, two of the Big-5 events, occurred immediately before and after the terrestrialization, respectively. It is noteworthy that these extinctions share common characteristics not only in extinction-relevant biological phenomena but also in non-biological aspects, suggesting possibly the same trigger/processes. Although the two major extinctions occurred during global cooling, the atmospheric compositions were significantly different between the Ordovician and Permian. The Ordovician $p\text{CO}_2$ (over 2000 ppm) was one order of magnitude higher than that of the Permian (ca. 300 ppm), and this suggests that the scenario of atmospheric CO_2 -driven global climate change is unlikely for causing these two extinction-related cases, and an alternative explanation is necessary

the mid-twentieth century, this scenario was not testable—i.e., it could not be “falsified” (by comparison to data) as required by Karl Popper’s definition of what constitutes science—and was therefore left ignored and almost forgotten.

Nearly 40 years later, by detecting a strong correlation between GCR flux into the atmosphere and satellite-observed cloud coverage in the lower atmosphere, Svensmark and Friis-Christensen (1997) proposed that global cooling/warming is controlled essentially by the cloud coverage of the globe, in response to GCR influx. Penetrating GCR can charge nitrate molecules in the atmosphere for coagulating vapor to form cloud particles (Svensmark et al. 2017). Two factors possibly related

with causes of Category 4 are of interest, i.e., the intensity of magnetic shield and that of GCR sources. One is earthbound, and the other is extraterrestrial/extrasolar.

Although the cosmoclimatological perspective *per se* is still under scrutiny, Isozaki (2009) applied this to a new interpretation on the Permian extinction case, by comparing the major change in the Earth's geomagnetism, i.e., the end of a long-term stable geomagnetic polarity (called the Kiaman Reverse Polarity Superchron; Fig. 18.4d) immediately before the G-LB extinction. The Earth's dipole geomagnetism, together with that of the Sun, forms an effective magnetic shield against the influx of GCR, and the change in shield strength along time controls the influx. Thus a major decrease in geomagnetic intensity is critical for allowing extensive cloud coverage to weaken or shut down solar irradiance (Fig. 18.7). Ancient episodes of large flux of GCR events may have invited past major cold spikes in human history (e.g., Usoskin et al. 2007). The G-LB and the end-Ordovician cooling events and relevant extinctions (Isozaki and Servais 2018) can be explained by this scenario rather than by the simple LIP/Traps-CO₂ story; nonetheless, mantle plume/LIPs may have been related to the modulation of geomagnetic intensity. The geomagnetic intensity is relevant to the convection pattern of molten metal in the outer core, which is related also to the convection of the mantle that includes mantle plume activities.

Moreover, multiple cooling events with glaciation in deep past may have been driven by the appearance of strong GCR sources in the neighbor of our Solar System. As mentioned above, in contrast, global cooling is indeed difficult to develop in the pre-Devonian Earth, if the greenhouse effect of atmospheric CO₂ alone was responsible for global surface temperature (Fig. 18.6). The major biodiversity drops called the Big-5 mass extinctions of the Phanerozoic (Sepkoski 1996; Alroy et al. 2008) can be correlated with multiple episodes of unusually increased GCR flux into the Earth's atmosphere (Svensmark and Calder 2007; Medvedev and Melott 2007). At least the end-Ordovician and G-LB extinctions occurred during unusual cooling intervals (Isozaki and Servais 2018), suggesting possible links to extraterrestrial events.

18.4.2 Past Chilling Events and Extinctions

Some *avant-garde* scientists further explored a possibility of episodic increase in GCR sources, such as supernovae, active galactic center, or dark cloud/nebula, and discussed the effect to the Solar System and the Earth's biosphere in the past (e.g., Svensmark and Friis-Christiansen 1997; Shaviv and Veizer 2003; Svensmark and Calder 2007; Medvedev and Melott 2007; Kataoka et al. 2014). Current astronomical observations revealed that the surrounding universe around our Solar System was not stable at all but highly changeable. For example, episodic starburst events occurred in the Milky Way Galaxy (Rocha-Pinto et al. 2000; de la Fuente Marcos and de la Fuente Marcos 2004). Starburst is a rare episode in the universe that creates numerous new stars in a limited time interval by the dynamic interactions

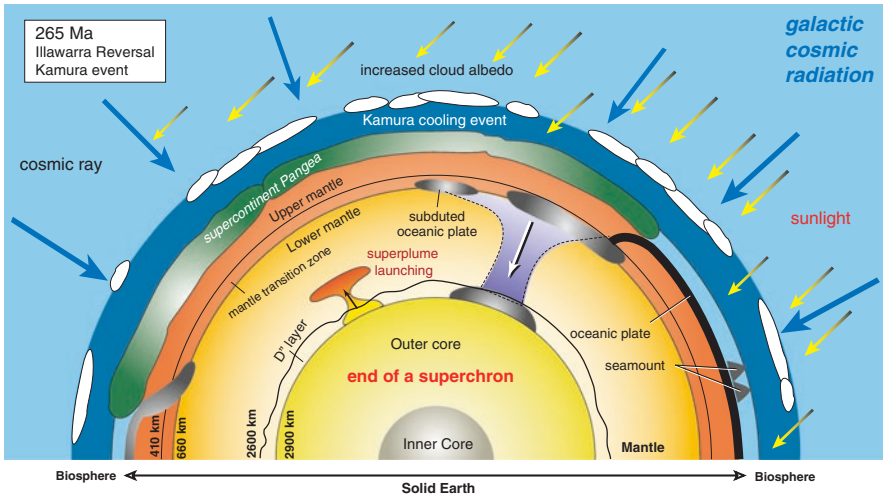


Fig. 18.7 Schematic diagram of the integrated “plume winter” scenario, showing a possible link among the flux of galactic cosmic radiation (GCR), global cloud coverage, global cooling, and mass extinction for explaining the unique G-LB event and other analogs (modified from Isozaki 2009b). GCR from extrasolar sources (e.g., supernovae, active galactic center, and/or dark cloud) can penetrate into the Earth’s atmosphere in large flux when the solar magnetic field (heliosphere) and geomagnetic field are not intense enough to shield the atmosphere. Large influx of GCR can charge nitrate molecules in the atmosphere for coagulating cloud particles. A cause-effect relationship between GCR flux and global coverage of lower cloud coupled with global cooling has been speculated (e.g., Svensmark and Friis-Christensen 1997; Svensmark and Calder 2007). Two factors determine the GCR flux into the atmosphere, i.e., the intensity of GCR source and the strength of the helio- and geomagnetic shield. The former is related to the relative position of our Solar System with respect to other galaxies. When our Solar System and/or our galaxy enters into a region in the universe with full of dark clouds (nebula), the Earth would receive much greater GCR flux than normal and may experience unusual global cooling. Recent astronomical observations preliminary detected major and minor starburst events in the past (de la Fuente Marcos and de la Fuente Marcos 2004). The latter factor is related to the long-term change in geomagnetism, whose intensity is controlled by the convection in the Earth’s outer core of the planet. When unusual conditions appear in the convection pattern, lowered geomagnetic intensity may allow more penetration of GCR than normal. The unique geomagnetic condition called superchrons, and its onset and cancelation, may correspond to such unusual conditions in geomagnetism. The exceptionally long-term Kiaman Reverse Polarity Superchron in the Late Carboniferous to Middle Permian (Irving and Parry 1963) and its collapse (called the Illawarra Reversal; Steiner 2006; Isozaki 2009a, b; Kirschvink et al. 2015) may be a harbinger of the Kamura global cooling event (Isozaki et al. 2007). In the late Middle Permian, a dramatic change likely took place just above the core/mantle boundary (in the D’ layer) owing to the drop of a relatively cold super-downwell (subducted oceanic slabs beneath Pangea) from the mantle transition zone (410–660 km deep). This could cause thermal instability on the core’s surface that can drive a major change in convection pattern of the outer core. The weakened geomagnetism likely allowed more GCR flux into the atmosphere to trigger extensive cloud coverage with higher albedo, which resulted in the Capitanian Kamura cooling event associated with the lowest sea level and selective decline/extinction of the unique tropical fauna. In summary, the flux of GCR likely controls the degree of cloud coverage over the Earth’s surface, which can determine global cooling/warming, more effectively than the claimed greenhouse effect of the atmospheric CO₂ and relevant H₂O (e.g., Svensmark and Friis-Christensen 1997; Shaviv and Veizer 2003). Major GCR events may have initiated past major cold spikes/glaciations and biodiversity loss (e.g., Usoskin et al. 2007; Medvedev and Melott 2007). In addition to the Big-5 extinctions of the Phanerozoic, two episodes of the dramatic snowball Earth (Sturtian and Marinoan episodes) in the Proterozoic, late Precambrian can be explained likewise (Svensmark and Calder 2007; Kataoka et al. 2014)

between galaxies, in particular by the collisions of plural galaxies. One of the consequences is the formation of numerous nebulae (dark clouds and supernova remnants); dark cloud is a dense, low-temperature interstellar cloud full of high-energy particles and submicrometer-sized dusts that block visible lights.

According to the recent astrophysical observations, starburst events occurred merely twice, in ca. 2.4–2.0 Ga and 0.8–0.6 Ga, in the history of Milky Way Galaxy, suggesting the frequent arrival of many dark clouds to our Solar System during the starburst intervals. Dark clouds/nebulae are capable of increasing GCR influx to the Earth and also shutting down the solar irradiance. Svensmark (2006) pointed out that the two snowball Earth events at 2.3 and 0.7 Ga likely correspond to the two significant episodes of GCR influx in timing.

Figure 18.8 illustrates possible correlation between starburst events in the solar neighborhood (de la Fuente Marcos and de la Fuente Marcos 2004) and significant catastrophes in the Earth's biosphere. Kataoka et al. (2014) pointed out that the two prominent peaks of starburst in the Proterozoic apparently coincide with the two snowball Earth events. The Neoproterozoic event with two major glaciations of the Sturtian and Marinoan episodes (Hoffman and Schrag 2002) occurred mostly during the 0.8–0.6 Ga starburst. Likewise, three minor peaks of starburst in the Phanerozoic occur slightly before the three major extinctions in the Paleozoic, i.e., the end-Ordovician, Late Devonian, and the G-LB/P-TB events. These correlations suggest that the intermittent arrival of dark clouds to our Solar System has caused significant environmental changes on the Earth's surface, in particular global cooling, although the time resolution for starbursts is not sufficiently high, and also certain time lags may have existed between the starburst peaks and arrival of dark clouds to our neighborhood.

For the Big-5 extinctions of the Phanerozoic (Fig. 18.1), no candidate of Category 4 cause has been presented to date. By analogy to the end-Ordovician and G-LB extinctions, the end-Devonian extinction may have occurred similarly during a global cooling/glaciation. The insignificant peak of starburst before the end-Triassic timing (Fig. 18.8) is concordant with the relatively minor degree of extinction (Stanley 2016; Lucas and Tanner 2018). The end-Cretaceous (K-PgB) bolide impact (Alvarez et al. 1980) may have been a minor side effect of gravitational disturbance of asteroids induced by an approaching dark cloud. Dark clouds are large-mass entities that may cause gravitational instability of asteroids/planets to induce falling of asteroid/meteorite onto the Earth's surface. Furthermore, larger/denser clouds may induce changes in the eccentricity of the Earth's inner core for modifying geodynamo and/or in crustal stress regime for triggering spontaneous magmatism.

The putative periodicity of extinction is in question (Erlykin et al. 2017); however, the long-term passage of our Solar System through multiple spiral arms of the Milky Way Galaxy one after another (Leitch and Vasisht 1998) may possibly have tuned the rhythm of GCR flux into the Earth and relevant biotic responses (Shaviv and Veizer 2003).

The GCR flux may dominantly control global climate, rather than the claimed greenhouse effect of the atmospheric CO₂ and relevant H₂O. Elevated flux of GCR may have caused the high degree of cloud coverage over the Earth's surface,

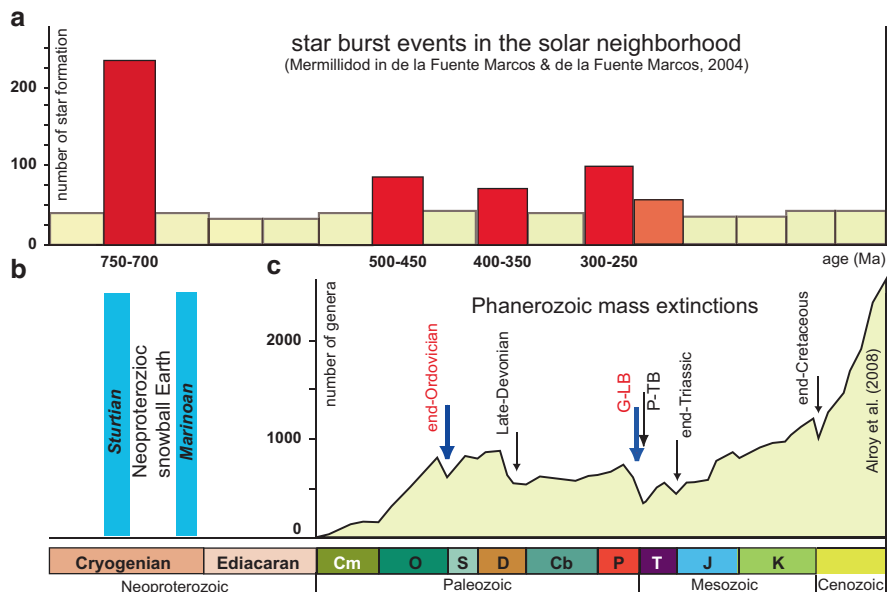


Fig. 18.8 Late Precambrian and Phanerozoic starburst events in the solar neighborhood and apparent coincidence in timing with the major mass extinction events in the Phanerozoic. **(a)** Secular change in star formation rate in the neighborhood of our Solar System during the last 800 myr in the Earth's history (after Mermilliod in de la Fuente Marcos and de la Fuente Marcos 2004). **(b)** Secular change in biodiversity during the Phanerozoic with the Big-5 mass extinctions (Alroy et al. 2008) and the Neoproterozoic snowball Earth event (Hoffman and Schrag 2002). Note the prominent peak of star formation (starburst event) around 0.7 Ga, which apparently coincides in timing with the Cryogenian snowball Earth event (Sturtian and Marinoan episodes). Likewise, three minor peaks in the Phanerozoic, although still distinct from the background level, occur slightly before the three major extinctions of the Paleozoic, i.e., the end-Ordovician, Late Devonian, and the G-LB/P-TB events. Although time resolution is still not sufficiently high, their apparent correlations suggest that the intermittent arrivals of dark clouds to our Solar System may have caused environmental changes on the Earth's surface, in particular global cooling/glaciations. Certain time lags may have existed between a starburst event and the arrival of dark clouds to the Solar System. The absence of a significant peak of star formation near the end-Triassic timing is concordant with the relatively minor degree of extinction (Lucas and Tanner 2018), whereas the end-Cretaceous (K-PgB) bolide impact (Alvarez et al. 1980) may have been a side effect of the gravitational disturbance of asteroids induced by approaching minor dark cloud(s)

controlling global cooling/warming. Not only a change in geomagnetic intensity of star/planet but also episodes of starburst in the universe may be responsible for increasing GCR, which can cause global environmental changes, in particular global cooling critical to majority of biota. Thus either internal and external forcing for the Earth's biosphere or both can be listed as ultimate cause(s) of mass extinction above all (Category 4; Table 18.1). Although many unknowns still remain, further study is definitely needed to prove or disprove the new scenarios. Given ultimate causes in cosmoclimatological forcing, more detailed explanations are inevitable

for relevant environmental changes (Category 2), various kill mechanisms (Category 1), and their mutual links for each major extinction event.

In short, past mass extinctions took place indeed on the Earth's surface; nonetheless the ultimate cause(s) might occur in the Earth's interior and/or its surroundings. Studies on mass extinction are currently entering a new generation with more synergic commitment with astrobiology. Unlike the mid-twentieth century, the new scenarios on the interaction with extraterrestrial agents became testable in the twenty-first century by virtue of innovative high-tech and high-resolution observations of extrasolar systems. On the other hand, modern geological/geochemical researches are under progress for checking whether or not we may detect materialistic lines of evidence for extraterrestrial/extrasolar flux from interstellar deep space, such as platinum group elements (PGEs), $^3\text{He}/^4\text{He}$ ratio, metallic compounds, etc.

18.5 Conclusions

This review summarized the current status of research on mass extinctions in the Phanerozoic, in particular the end-Paleozoic extinction that determined the fate of modern animals including mammals. The following conclusions were reached.

1. The largest mass extinction of the Phanerozoic at the end-Permian in fact occurred in two steps, i.e., first at and around the G-LB (ca. 259 Ma) and second at the P-TB (ca. 252 Ma) (Figs. 18.1 and 18.2). The biodiversity loss of the former marked the first major decline of the long-lived Carboniferous-Permian fauna, and it occurred in the very unique interval of various aspects, in particular during the global cooling with the lowest sea level of the Phanerozoic.
2. Extinction causes are classified into four categories from small to large scale: Category 1, direct killing mechanism; Category 2, global background environmental change; Category 3, trigger appearing on the Earth's surface; and Category 4, ultimate cause (Table 18.1). Among all, causes of the Category 4 are the most significant thus critical to induce other causes of lower categories in cascade.
3. The G-LB and end-Ordovician extinctions share more similarities, in particular global cooling (Category 2), than differences (Fig. 18.4), suggesting that both extinctions may have been triggered by the same cause/processes. Nonetheless the background atmospheric compositions were totally different between them: amount of CO_2 before and after the mid-Paleozoic land terrestrialization. An alternative explanation is needed for the common global cooling instead of the greenhouse effect of the atmosphere.
4. Besides the currently most prevalent scenario of the Traps/LIPs (Category 3)-extinction link, a new cosmoclimatological cause is proposed with respect to astrobiology. The increase of GCR influx to the Earth's atmosphere can drive extensive global cloud coverage/blocking solar irradiance, which can induce global cooling and resulting extinction (Fig. 18.7). Influx of GCR is controlled

both by external and internal forcing, i.e., the intensity of the sources from the extrasolar space and the strength of geomagnetic shield formed in the Earth's core. Influx of GCR episodically increased by the arrival of dark clouds to the neighborhood of the Solar System with respect to the contemporary strength of geomagnetic shield. The Big-5 extinctions of the Phanerozoic (Fig. 18.1) and the Neoproterozoic snowball Earth event (Fig. 18.8) can be explained consistently by the same ultimate cause (Category 4).

Acknowledgments Prof. Robert Geller (Univ. Tokyo) provided valuable comments on the manuscript. Tomoyo Tobita and Hikaru Sawada helped in drafting. This study was funded by KAKENHI (grant-in-aid from the Japan Society for Promotion of Science; no. 26257212).

References

- Algeo TJ, Kuwahara K, Sano H, Bates S, Lyons T, Elswick E, Hinnov L, Ellwood B, Moser J, Maynard JB (2011) Spatial variation in sediment fluxes, redox conditions, and productivity in the Permian–Triassic Panthalassic Ocean. *Palaeogeogr Palaeoclimatol Palaeoecol* 308:65–83
- Alroy J, Aberhan M, Bottjer DJ et al (2008) Phanerozoic trends in the global diversity of marine invertebrates. *Science* 321:97–100
- Alvarez LW, Alvarez W, Asaro F, Michel HV (1980) Extraterrestrial cause for the Cretaceous–Tertiary extinction: experimental results and theoretical interpretation. *Science* 208:1095–1108
- Bambach RK (2006) Phanerozoic biodiversity mass extinctions. *Ann Rev Earth Planet Sci* 34:127–155
- Benton MJ, Newell AJ (2014) Impacts of global warming on Permo-Triassic terrestrial ecosystems. *Gondwana Res* 25:1308–1337
- Baumiller TK, Messing CG (2007) Stalked crinoid locomotion, and its ecological and evolutionary implications. *Palaeontol Electron* 10:1–10
- Becker L, Podera R, Hunt AG, Bunch TE, Rampino M (2001) Impact event at the Permian–Triassic boundary: evidence from extraterrestrial noble gases in fullerenes. *Science* 291:1530–1533
- Belica ME, Tohver E, Pisarevsky SA, Jourdan F, Denyszyn S, George AD (2017) Middle Permian paleomagnetism of the Sydney Basin, eastern Gondwana: testing Pangea models and the timing of the end of the Kiaman Reverse Superchron. *Tectonophysics* 699:178–198
- Berner RA (2006) GEOCARBSULF: a combined model for Phanerozoic atmospheric O₂ and CO₂. *Geochim Cosmochim Acta* 70:5653–5664
- Bond DPG, Grasby SE (2017) On the causes of mass extinctions. *Palaeogeogr Palaeoclimatol Palaeoecol* 478:3–29
- Bond DPG, Hilton J, Wignall PB, Ali JR, Stevens LG, Sun YD, Lai XL (2010) The Middle Permian (Capitanian) mass extinction on land and in the oceans. *Earth Sci Rev* 102:100–116
- Burgess SD, Bowring SA, Shen SZ (2014) High-precision timeline for Earth's most severe extinction. *Proc Natl Acad Sci USA* 111:3316–3321
- Burgess SD, Muirhead JD, Bowring SA (2017) Initial pulse of Siberian Traps sills as the trigger of the end-Permian mass extinction. *Nat Commun* 8:164
- Campbell I, Czamanske GK, Fedorenko VA, Hill RI, Stepanov V (1992) Synchronism of the Siberian traps and the Permian–Triassic boundary. *Science* 258:1760–1763
- Cao CQ, Love GD, Hays LE, Wang W, Shen SZ, Summons RE (2009) Biogeochemical evidence for euxinic oceans and ecological disturbance presaging the end-Permian mass extinction event. *Earth Planet Sci Lett* 281:188–201

- Chung SL, Jahn BM, Wu GY, Lo CH, Cong BL (1998) The Emeishan flood basalt in SW China: a mantle plume initiation model and its connection with continental breakup and mass extinction at the Permian–Triassic boundary. *Am Geophys Union Geodyn Ser* 27:47–58
- Clapham ME, Shen SZ, Bottjer DJ (2009) The double mass extinction revisited: reassessing the severity, selectivity, and causes of the end-Guadalupian biotic crisis (Late Permian). *Palaeobiology* 35:32–50
- Courtillot V (1999) *Evolutionary catastrophes: the science of mass extinctions*. Cambridge University Press, Cambridge, UK
- de la Fuente Marcos R, de la Fuente Marcos C (2004) On the correlation between the recent star formation rate in the solar neighbourhood and the glaciations period record on Earth. *New Astron* 10:53–66
- Eldredge N, Gould SJ (1972) Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf TJM (ed) *Models in paleobiology*. Freeman Cooper, San Francisco, pp 82–115
- Erylkin AD, Harper DAT, Sloan T, Wolfendale AW (2017) Mass extinctions over the last 500 Myr: an astronomical cause? *Palaeontology* 60:159–167
- Ernst RE (2014) *Large igneous provinces*. Cambridge University Press, Cambridge, UK
- Erwin DH (2006) *Extinction*. Princeton University Press, New York
- Farley KA, Mukhopadhyay S (2001) An extraterrestrial impact at the Permian–Triassic boundary? *Science* 293:U1–U3
- Farley KA, Ward P, Garrison G, Mukhopadhyay S (2005) Absence of extraterrestrial ^3He in Permian–Triassic age sedimentary rocks. *Earth Planet Sci Lett* 240:265–275
- Federenko V, Czamanske G, Zen'ko T, Budhan D (2000) Field and geochemical studies of the melilite-bearing Arydhangsky Suite, and overall perspective on the Siberian alkaline-ultramafic flood-volcanic rocks. *Int Geol Rev* 42:769–804
- Fielding CR, Frank TD, Birgenheier LP, Rygel MC, Jones AT, Roberts J (2008) Stratigraphic imprint of the Late Palaeozoic ice age in eastern Australia: a record of alternating glacial and nonglacial climate regime. *J Geol Soc Lond* 165:129–140
- Grasby SE, Sanei H, Beauchamp B (2011) Catastrophic dispersion of coal fly ash into oceans during the latest Permian extinction. *Nat Geosci* 4:104–107
- Grasby SE, Beauchamp B, Bond DPG, Wignall PB, Sanei H (2015) Mercury anomaly associated with three extinction events (Capitanian crisis, Latest Permian Extinction, and the Smithian/Spathian Extinction) in NW Pangea. *Geol Mag* 153:285–297
- Grice K, Cao C, Love GD, Bottcher ME, Twitchett R, Grosjean E, Summons RE, Turgeon SC, Dunning W, Jin YG (2005) Photic zone euxinia during the Permian–Triassic superanoxic event. *Science* 307:706–709
- Hallam, A., Wignall, P.B., 1997. *Mass extinctions and their aftermath*. Oxford University Press, Oxford
- Haq BU, Schutter SR (2008) A chronology of Paleozoic sea-level changes. *Science* 322:64–68
- Hoffman PF, Schrag DP (2002) The snowball Earth hypothesis: testing the limits of global change. *Terra Nova* 14:129–155
- Holser WT, Magaritz M (1987) Events near the Permian–Triassic boundary. *Mod Geol* 11:155–180
- Irving E, Parry LG (1963) The magnetism of some Permian rocks from New South Wales. *Geophys J R Astron Soc* 7:395–411
- Isozaki Y (1997) Permo–Triassic boundary superanoxia and stratified superocean: records from lost deep-sea. *Science* 276:235–238
- Isozaki Y (2001) An extraterrestrial impact at the Permian–Triassic boundary? *Science* 293:U1–U3
- Isozaki Y (2009a) The Illawarra Reversal: a fingerprint of the superplume triggering Pangean breakup and end-Guadalupian (Permian) extinction. *Gondwana Res* 15:421–432
- Isozaki Y (2009b) Integrated plume winter scenario for the double-phased extinction during the Paleozoic–Mesozoic transition: G-LB and P-TB events from a Panthalassan perspective. *J Asian Earth Sci* 36:459–480
- Isozaki Y (2014) Memories of pre-Jurassic lost oceans: how to retrieve them from extant lands. *Geosci Can* 41:283–311

- Isozaki Y, Aljinović D (2009) End-Guadalupian extinction of the Permian gigantic bivalve Alatoconchidae: end of gigantism in tropical seas by cooling. *Palaeogeogr Palaeoclimatol Palaeoecol* 284:11–21
- Isozaki Y, Servais T (2018) The Hirnantian (Late Ordovician) and end-Guadalupian (Middle Permian) mass extinction events compared. *Lethaia* 51:173–186 in press
- Isozaki Y, Kawahata H, Ota A (2007) A unique carbon isotope record across the Guadalupian-Lopingian (Middle-Upper Permian) boundary in mid-oceanic paleoatoll carbonates: the high-productivity “Kamura event” and its collapse in Panthalassa. *Glob Planet Chang* 55:21–38
- Isozaki Y, Aljinovic D, Kawahata H (2011) The Guadalupian (Permian) Kamura event in European Tethys. *Palaeogeogr Palaeoclimatol Palaeoecol* 308:12–21
- Jin YG, Zhang J, Shang QH (1994) Two phases of the end-Permian mass extinction. *Can Soc Petrol Geol Mem* 17:813–822
- Kaiho K, Kajiura Y, Nakano T, Miura Y, Kawahata H, Tazaki K, Ueshima M, Chen ZQ, Shi GR (2001) End-Permian catastrophe by a bolide impact: evidence of a gigantic release of sulfur from the mantle. *Geology* 29:815–818
- Kamo SL, Czamanske GK, Krogh TE (1996) A minimum U-Pb age for Siberian flood basalt volcanism. *Geochim Cosmochim Acta* 60:3505–3511
- Kani T, Hisanabe C, Isozaki Y (2013) The Capitanian minimum of $^{87}\text{Sr}/^{86}\text{Sr}$ ratio in the Permian mid-Panthalassan paleo-atoll carbonates and its demise by the deglaciation and continental doming. *Gondwana Res* 24:212–221
- Kataoka R, Ebisuzaki T, Miyahara H, Nimura T, Tomida T, Sato T, Maruyama S (2014) The Nebula Winter: the united view of the snowball Earth, mass extinctions, and explosive evolution in the late Neoproterozoic and Cambrian periods. *Gondwana Res* 25:1153–1163
- Kiessling W (2001) Paleoclimatic significance of Phanerozoic reefs. *Geology* 29:751–754
- Kirschvink JL, Isozaki Y, Shibuya H, Otofujii Y, Raub TD, Hilburn IA, Kasuya T, Yokoyama M, Bonifacie M (2015) Challenging the sensitivity limits of paleomagnetism: magnetostratigraphy of weakly magnetized Guadalupian-Lopingian (Permian) limestone from Kyushu. *Jpn Palaeogeogr Palaeoclimatol Palaeoecol* 418:75–89
- Knoll AH, Bambach RK, Canfield DE, Grotzinger JP (1996) Comparative Earth history and Late Permian mass extinction. *Science* 273:452–457
- Koeberl C, Gilmour I, Reimold WU, Claeys P, Ivanov B (2002) End-Permian catastrophe by bolide impact: evidence of a gigantic release of sulfur from the mantle: comment and reply. *Geology* 30:855–856
- Kofukuda D, Isozaki Y, Igo H (2014) A remarkable sea-level drop across the Guadalupian-Lopingian (Permian) boundary in low-latitude mid-Panthalassa: irreversible changes recorded in accreted paleo-atoll limestones in Akasaka and Ishiyama. *Jpn J Asian Earth Sci* 82:47–65
- Korte C, Jasper T, Kozur HW, Veizer J (2005) $\delta^{18}\text{O}_{\text{carb}}$ and $\delta^{13}\text{C}_{\text{carb}}$ of Permian brachiopods: a record of seawater evolution and continental glaciation. *Palaeogeogr Palaeoclimatol Palaeoecol* 224:333–351
- Korte C, Jasper T, Kozur HW, Veizer J (2006) $^{87}\text{Sr}/^{86}\text{Sr}$ record of Permian seawater. *Palaeogeogr Palaeoclimatol Palaeoecol* 240:89–107
- Kump L, Arthur MA (1999) Interpreting carbon-isotope excursions: carbonates and organic matter. *Chem Geol* 161:181–198
- Leitch EM, Vasist G (1998) Mass extinctions and the sun’s encounters with spiral arms. *New Astron* 3:51–56
- Looy CV, Twitchett RJ, Dilcher DL, Van Konijnenburg-Van Cittert JHA, Visscher H (2001) Life in the end-Permian dead zone. *Proc Natl Acad Sci U S A* 98:7879–7883
- Lucas SG (2009) Timing and magnitude of tetrapod extinctions across the Permo-Triassic boundary. *J Asian Earth Sci* 36:491–502
- Lucas SG, Shen SZ (2018) The Permian timescale. *Geol Soc Lond Spec Publ* 450:1–19
- Lucas SG, Tanner LH (2018) The missing mass extinction at the Triassic-Jurassic boundary. In: Tanner LH (ed) *The Late Triassic world*. Topics in geobiology, vol 46. Springer, Heidelberg, pp 721–785

- Matsuda T, Isozaki Y (1991) Well-documented travel history of Mesozoic pelagic chert in Japan: from remote ocean to subduction zone. *Tectonics* 10:475–499
- McArthur JM, Howarth RJ, Shields GA (2012) Strontium isotope stratigraphy. In: Gradstein FM, Ogg JG, Schmitz MD, Ogg GM (eds) *Geologic time scale 2012*. Elsevier, Amsterdam, pp 127–144
- Medvedev MV, Melott AL (2007) Do extragalactic cosmic rays induce cycles in fossil diversity? *Astrophys J* 664:879–889
- Musashi M, Isozaki Y, Koike T, Kreulen R (2001) Stable carbon isotope signature in mid-Panthalassa shallow-water carbonates across the Permo-Triassic boundary: evidence for ¹³C-depleted ocean. *Earth Planet Sci Lett* 196:9–20
- Nabbefeld B, Grice K, Summons R, Hays L, Cao C (2010) Significance of polycyclic aromatic hydrocarbons (PAHs) in Permian/Triassic boundary sections. *Appl Geochem* 25:1374–1382
- Paytan A, Gray ET (2012) Sulfur isotope stratigraphy. In: Gradstein FM, Ogg JG, Schmitz MD, Ogg GM (eds) *Geologic time scale 2012*. Elsevier, Amsterdam, pp 167–180
- Raup DM, Sepkoski JJ (1982) Mass extinctions in the marine fossil record. *Science* 215:1501–1503
- Reichow MK, Pringle MS, Al'Mukhamedov AI, Allen MB, Andreichev VL, Buslov MM, Davies CE, Fedoseev GS, Fitton JG, Inger S, Medvedev AY, Mitchell C, Puchkov VN, Safonova IY, Scott RA, Saunders AD (2009) The timing and extent of the eruption of the Siberian Traps large igneous province: implications for the end-Permian environmental crisis. *Earth Planet Sci Lett* 277:9–20
- Renne PR, Basu AR (1991) Rapid eruption of the Siberian Traps flood blasts at the Permian–Triassic boundary. *Science* 253:176–179
- Retallack GJ (2013) Permian and Triassic greenhouse crises. *Gondwana Res* 24:90–103
- Rocha-Pinto HJ, Scalo J, Maciel WJ, Flynn C (2000) Chemical enrichment and star formation in the Milky Way disk II. *Star Format Hist Astron Astrophys* 358:869–885
- Royer DL, Donnadieu Y, Park J, Kowalczyk J, Godderis Y (2014) Error analysis of CO₂ and O₂ estimates from the long-term geochemical model GEOCARBSULF. *Am J Sci* 314:1259–1283
- Rubidge BS, Erwin DH, Ramezani J, Bowring SA, De Klerk WJ (2013) High-precision temporal calibration of late Permian vertebrate biostratigraphy: U–Pb zircon constraints from the Karoo Supergroup. *S Afr Geol* 41:363–366
- Saitoh M, Isozaki Y, Yao JX, Ji ZS, Ueno Y, Yoshida H (2013) Lithostratigraphy across the Guadalupian-Lopingian (Middle-Upper Permian) boundary at Chaotian in Sichuan, South China: secular change in sea level and redox condition. *Glob Planet Chang* 105:180–192
- Saitoh M, Ueno Y, Isozaki Y, Nishizawa M, Shozugawa K, Kawamura T, Yao JX, Ji ZS, Takai K, Yoshida H, Matsuo M (2014) Isotopic evidence for water-column denitrification and sulfate reduction preceding the end-Guadalupian (Permian) extinction. *Glob Planet Chang* 123:110–120
- Saltzman MR, Thomas E (2012) Carbon isotope stratigraphy. In: Gradstein FM, Ogg JG, Schmitz MD, Ogg GM (eds) *Geologic time scale 2012*. Elsevier, Amsterdam, pp 207–232
- Saunders AD, England RW, Reichow MK, White RV (2005) A mantle plume origin for the Siberian traps: uplift and extension in the West Siberian Basin. *Russ Lithos* 79:407–424
- Schindewolf OH (1955) Ueber die möglichen Ursachen der grossen erdgeschichtlichen Faunenschnitte. *Neus Jahrb Palaontol* 1954:457–465
- Scotese CR (2008) PALEOMAP project. <http://www.scotese.com>
- Sepkoski JJ Jr (1996) Patterns of Phanerozoic extinction: a perspective from global data bases. In: Walliser OH (ed) *Global events and event stratigraphy in the Phanerozoic*. Springer, Heidelberg, pp 35–51
- Shaviv NJ, Veizer J (2003) Celestial driver of Phanerozoic climate? *GSA Today* 13:4–10
- Shellnutt JG (2014) The Emeishan large igneous province: a synthesis. *Geosci Front* 5:369–394
- Shen SZ, Shi GR (2002) Paleobiogeographical extinction patterns of Permian brachiopods in the Asian-Western Pacific region. *Paleobiology* 28:449–463
- Shen YN, Farquhar J, Zhang H, Masterson A, Zhang TG, Wing BA (2011) Multiple S-isotopic evidence for episodic shoaling of anoxic water during Late Permian mass extinction. *Nat Commun* 2:210

- Shen SZ, Cao C, Zhang H, Bowring SA, Henderson CM, Payne JL, Davydov VI, Chen B, Yuan D, Zhang Y, Wang W, Zhang Q (2013) High-resolution $\delta^{13}\text{C}_{\text{carb}}$ chemostratigraphy from latest Guadalupian through earliest Triassic in South China and Iran. *Earth Planet Sci Lett* 37:156–165
- Shi L, Feng QL, Shen J, Ito T, Chen ZQ (2016) Proliferation of shallow-water radiolarians coinciding with enhanced oceanic productivity in reducing conditions during the Middle Permian, South China: evidence from the Gufeng Formation of western Hubei Province. *Palaeogeogr Paleoclimatol Palaeoecol* 444:1–14
- Stanley SM (1988) Climate cooling and mass extinction of Paleozoic reef communities. *PALAIOS* 3:228–232
- Stanley SM (2016) Estimates of the magnitudes of major marine mass extinctions in earth history. *Proc Natl Acad Sci U S A* 113:E6325–E6334
- Stanley SM, Yang X (1994) A double mass extinction at the end of the Paleozoic era. *Science* 266:1340–1344
- Steiner MB (2006) The magnetic polarity time scale across the Permian–Triassic boundary. *Geol Soc Lond Spec Publ* 265:15–38
- Sun Y, Joachimski MM, Wignall PB, Yan C, Chen Y, Jiang H, Wang L, Lai X (2012) Lethally hot temperatures during the early Triassic greenhouse. *Science* 338:366–370
- Svensen H, Planke S, Polozov AG, Schmidbauer N, Corfu F, Podladchikov YY, Jamtveit B (2009) Siberian gas venting and the end-Permian environmental crisis. *Earth Planet Sci Lett* 277:490–500
- Svensmark H (2006) Cosmic rays and biosphere over 4 billion years. *Astron Nachr* 327:871–875
- Svensmark H, Calder N (2007) *The chilling stars: a new theory of climate change*. Icon Books, London
- Svensmark H, Friis-Christensen E (1997) Variation of cosmic ray flux and global cloud coverage – a missing link in solar-climate relationships. *J Atmos Solar Terr Phys* 59:1225–1232
- Svensmark H, Enghoff MB, Shaviv NJ, Svensmark J (2017) Increased ionization supports growth of aerosols into cloud condensation nuclei. *Nat Commun* 8:2199
- Usoskin IG, Solanki SK, Kovaltsov GA (2007) Grand minima and maxima of solar activity: new observational constraints. *Astron Astrophys* 471:4301–4309
- Veizer J, Ala D, Azmy K, Bruckschen P, Buhl D, Bruhm F, Carden GAF, Diener A, Ebner S, Godderis Y, Jasper T, Korte C, Pawellek F, Podlaha OG, Strauss H (1999) $^{87}\text{Sr}/^{86}\text{Sr}$, $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{18}\text{O}_{\text{carb}}$ evolution of Phanerozoic seawater. *Chem Geol* 161:59–88
- Wang XD, Sugiyama T (2000) Diversity and extinction patterns of Permian coral faunas of China. *Lethaia* 33:285–294
- Wang W, Cao CQ, Wang Y (2004) The carbon isotope excursion on GSSP candidate section of Lopingian-Guadalupian boundary. *Earth Planet Sci Lett* 220:57–67
- Whiteside JH, Grice K (2016) Biomarker records associated with mass extinction events. *Annu Rev Earth Planet Sci* 44:581–612
- Wignall PB (2001) Large igneous provinces and mass extinctions. *Earth Sci Rev* 53:1–33
- Wignall PB, Hallam A (1992) Anoxia as a cause of the Permian/Triassic mass extinction: facies evidence from northern Italy and western United States. *Palaeogeogr Palaeoclimatol Palaeoecol* 93:21–46
- Wignall PB, Sun YB, Bond DPG, Izon G, Newton RJ, Vedrine S, Widdowson M, Ali JR, Lai XL, Jian HS, Cope H, Botrell SH (2009) Volcanism, mass extinction, and carbon isotope fluctuations in the Middle Permian of China. *Science* 324:1179–1182
- Xie S, Pancost RD, Huang J, Wignall PB, Yu JX, Tang XY, Chen L, Huang XY, Lai XL (2007) Changes in the global carbon cycle occurred as two episodes during the Permian–Triassic crisis. *Geology* 35:1083–1086
- Zhong YT, He B, Mundil R, Xu YG (2014) CA-TIMS zircon U–Pb dating of felsic ignimbrite from the Binchuan section: implications for the termination age of Emeishan large igneous province. *Lithos* 204:14–19

Chapter 19

Mass Extinction at the Cretaceous–Paleogene (K–Pg) Boundary



Teruyuki Maruoka

Abstract One of the “Big Five” mass extinctions in the Phanerozoic Eon occurred at the Cretaceous–Paleogene (K–Pg) boundary (66.0 million years ago). The K–Pg mass extinction was triggered by a meteorite impact that produced a crater at Chicxulub on the Yucatán Peninsula, Mexico. The following environmental perturbations might have been induced by the Chicxulub impact and acted as the killing mechanisms for the K–Pg mass extinction: (1) sunlight shielding, (2) sulfuric and nitric acid rain, (3) CO₂-induced global warming, (4) ultraviolet penetration, and (5) toxic effects of ground-level ozone. The details of these perturbation events are summarized in this chapter. Based on evidence in sedimentary rocks, we could confirm whether such perturbation events occurred or not. However, it was difficult to reconstruct quantitatively the magnitudes and durations for such perturbation events because the necessary time-resolved information (yearly to millennium-scale) is lacking.

Keywords K–Pg mass extinction · Meteorite impact · Sunlight shielding · Acid rain

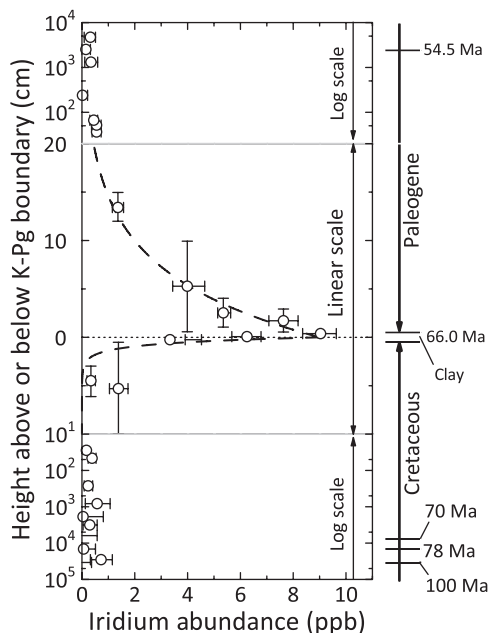
19.1 Introduction

One of the “Big Five” mass extinctions during the Phanerozoic Eon occurred at the Mesozoic and Cenozoic [Cretaceous–Paleogene (K–Pg)] boundary (66.0 million years ago). It has been estimated that about 40% of genera (Sepkoski 1996) and 70% of species went extinct at the K–Pg boundary, and it ranked third among the Phanerozoic mass extinctions (Stanley 2016). The Mesozoic species were replaced by Cenozoic species at the boundary clay layer, in which Alvarez et al. (1980) found an anomalously high concentration of iridium (Fig. 19.1). As iridium is one of the siderophile elements that favorably concentrate in metallic iron, it should have been

T. Maruoka (✉)

Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: maruoka.teruyuki.fu@u.tsukuba.ac.jp

Fig. 19.1 Iridium abundances for Italian limestones around the Cretaceous–Paleogene boundary (Data are from Alvarez et al. 1980). Dashed curves represent exponential fitting with half-heights of 4.6 cm and 0.43 cm, respectively



largely eliminated from the Earth's surface during the formation of the metallic core. In order to explain such iridium enrichment, the authors proposed that such iridium might have been supplied by a meteorite or meteorites because such objects have higher concentrations of iridium than crust materials. In addition, Ganapathy (1980) reported the enrichment of not only iridium but also other siderophile elements in the boundary clays (Fig. 19.2). The elemental patterns of siderophile elements in the clays were very similar to those of meteorites, but the patterns were different from those of crustal and mantle-derived materials.

After Alvarez et al. (1980) proposed the impact hypothesis, the occurrence of an impact was demonstrated by the recognition of impact-related materials in the boundary clays, including shocked quartz (Bohor et al. 1984, 1986, 1987), Ni-rich spinels (Bohor et al. 1986; Kyte and Smit 1986), stishovite (high-pressure polymorph of SiO_2) (McHone et al. 1989), microdiamonds (Carlisle and Braman 1991; Gilmour et al. 1992), and impact glass spherules (microtektites) (Izett 1991; Sigurdsson et al. 1991). The distribution of such ejecta points to an impact event in the Gulf of Mexico–Caribbean region. Finally, a 200-km-size crater structure was found at Chicxulub on the Yucatán Peninsula (Fig. 19.3; Hildebrand et al. 1991). The $^{40}\text{Ar}/^{39}\text{Ar}$ age of the glassy melt rock recovered from drill core samples of the Chicxulub crater is indistinguishable from that obtained for microtektites in the K–Pg boundary clays (Swisher et al. 1992). The temporal match between the ejecta layer and the onset of the extinctions leads us to conclude that the Chicxulub impact triggered the K–Pg mass extinction (Schulte et al. 2010). The eruption of large volumes of lavas in the Deccan Traps overlapped the K–Pg boundary and has been

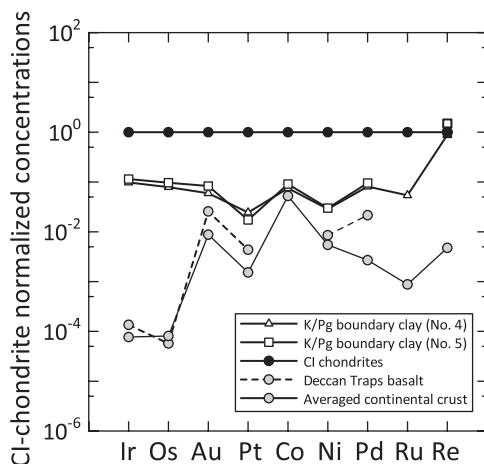


Fig. 19.2 Siderophile element abundances in various rock samples normalized to those of CI chondrites. Data are from Ganapathy (1980) for K–Pg boundary clays, Crocket and Paul (2004) for Deccan Traps basalt, Rudnick and Gao (2003) for the average values of the continental crust, and Palme and Jones (2003) for CI chondrites. Data are from Ganapathy (1980) for two specimens (NO. 4 and 5) of K–Pg boundary clays at Stevns Klint in Denmark. The elemental patterns of siderophile elements, except for Re, for the K–Pg boundary clays are relatively flat, whereas those for the Deccan Traps and continental crust are not

cited as a trigger of the extinctions. However, the main eruptive phase of the Deccan Traps preceded the K–Pg boundary (Robinson et al. 2009; Schoene et al. 2015). In this chapter, the environmental perturbations induced by the meteorite impact are briefly reviewed.

19.2 Sunlight Shielding by Submicron-Size Grains

19.2.1 Clastic Dusts

Alvarez et al. (1980) proposed a mechanism that could lead to a mass extinction after a meteorite impact. First, a meteorite impact would produce a dust cloud that could potentially reach the stratosphere, and then, submicron-size dust particles would remain in the stratosphere for a few years or so, which would shield the sunlight. Such sunlight shielding would then induce the shutdown of photosynthesis and lead to the destruction of ecosystems. Later research revealed that 10^{16} g of submicron-size grains are needed to shut down photosynthesis (e.g., Toon et al. 1982). However, the mass of submicron-size clastic dusts was estimated to be less than 10^{14} g, which is at least two orders of magnitude less than that needed to shut down photosynthesis (Pope 2002). Therefore, sunlight shielding by the clastic dust produced by the K–Pg impact would not have shut down photosynthesis.

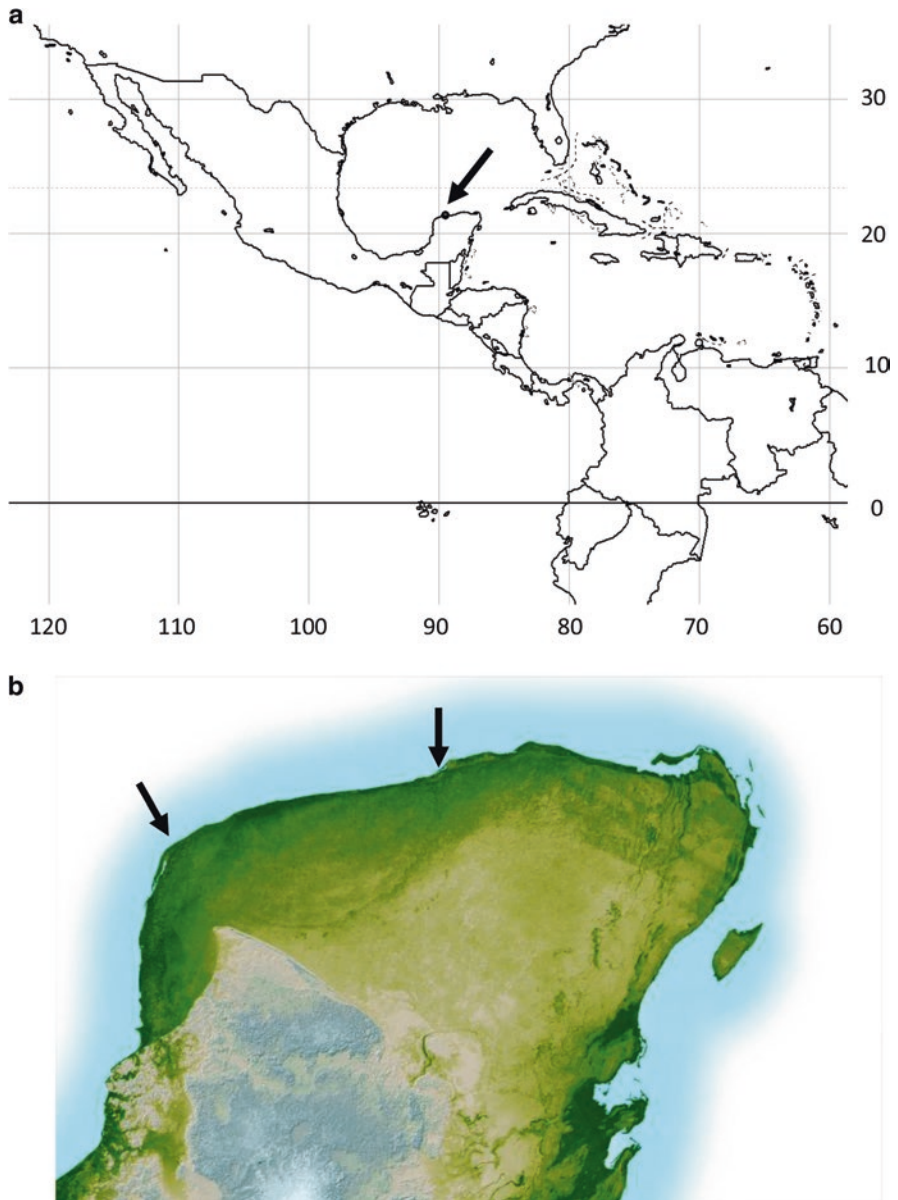


Fig. 19.3 (a) Location of the Chicxulub crater on the Yucatán Peninsula, Mexico, and (b) its radar image from the Space Shuttle Endeavour (Courtesy NASA/JPL-Caltech). As the topography in this image has been exaggerated, the trough produced by the meteorite impact (the darker arching curve in the upper left corner of the peninsula between the two arrows) can be seen. This trough is a part of 180-km-diameter circle

19.2.2 Soot

Wolbach et al. (1985, 1990) documented an enrichment of soot in the K–Pg boundary clays and proposed global fires as its cause. The global amount of soot was estimated to be $\sim 7 \times 10^{16}$ g (Wolbach et al. 1990), which would have been sufficient to induce sunlight shielding if it was supplied to the stratosphere over a short period of time. Melosh et al. (1990) proposed that the thermal radiation produced by the ballistic reentry of ejecta may have been responsible for the ignition of global wildfires.

As soot from one large fire could spread globally, the occurrence of soot over the whole Earth may not necessarily reflect the occurrence of global wildfires. In addition, soot can be produced from a variety of sources such as the combustion of coal, oil, gas, and biomass. On the other hand, charcoal is only formed through the combustion of biomass. Therefore, enrichment of charcoal should have spread globally if global wildfires occurred at the K–Pg boundary. However, Belcher et al. (2003, 2005) found the charcoal concentrations to be below Late Cretaceous background levels in the terrestrial K–Pg boundary clays of North America, thus suggesting that global wildfires might not have occurred at the K–Pg impact. Instead of global fires, Belcher et al. (2003, 2005) proposed that vaporization of hydrocarbons from the impact site served as a source for soot in the K–Pg boundary clays. However, the hydrocarbon abundance in the target rocks is far too low to explain the observed amount of soot in the K–Pg boundary clays (Robertson et al. 2004, 2013; Morgan et al. 2013). Robertson et al. (2004) suggested that the observed absence of charcoal might be evidence for a fire of unusually high intensity. An intense fire would lose less heat because of the lower surface-to-volume ratio, which could have led to temperatures high enough to destroy charcoal.

In the charcoal from the K–Pg boundary, Jones and Lim (2000) found features resulting from biodegradation before carbonization. This observation implies that there could have been a significant time lag (month, years, and possibly decades) between plant death and fires. Therefore, such fires might have not immediately followed the K–Pg impact; instead, they could have been induced by global warming following CO₂ releases after the K–Pg impact. This implies that soot production might have been prolonged, which could have reduced the intensity of darkness induced by soot accumulation in the stratosphere. Overall, the distribution and timing of wildfires at the K–Pg impact, and therefore, their influences on sunlight shielding are still poorly understood (see Belcher et al. 2009, 2015; Kaiho et al. 2017).

19.2.3 Sulfate Aerosols

Sulfur-containing gases (SO₂ or SO₃) were released from the anhydrite-containing target site of the K–Pg impact (evaporite layer in Fig. 19.4; López-Ramos 1975). Such gases would have been injected into the stratosphere and converted to sulfuric

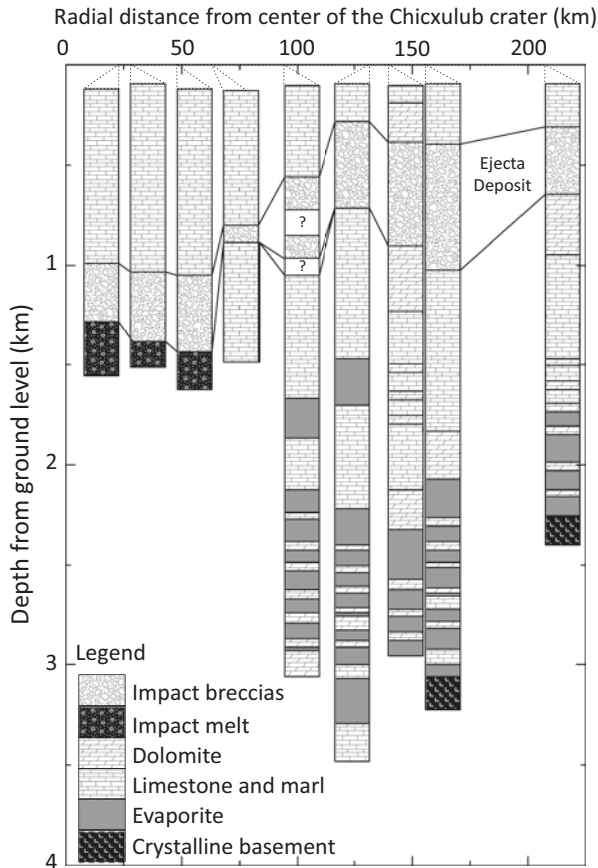


Fig. 19.4 Schematic columns of the boreholes obtained by Petr leos Mexicanos and the Yaxcopoil-1 borehole (Stinnesbeck et al. 2004), showing the main lithological divisions (Reproduced from Urrutia-Fucugauchi et al. 2011). Evaporite in the boreholes mainly consists of anhydrite (Ward et al. 1995)

acid aerosols, which could have caused considerable cooling of the Earth's surface (Brett 1992; Sigurdsson et al. 1992; Pierazzo et al. 1998). The aerosols would then have fallen back to the surface as acid rain. The amount of sulfur released from the target is estimated to have been 45–75 and 76–127 Gt ($= 10^{15}$ g) of S for asteroids of 10 and 15 km in diameter, respectively, when they hit the Earth at a speed of 20 km/s (Table 19.1; Pierazzo et al. 1998). Sulfur in the projectile also might have been supplied. A carbonaceous chondritic projectile of 10 and 15 km in diameter would have contained 33–42 and 82–106 Gt of S, respectively (Table 19.1; Maruoka and Koeberl 2003).

As the sulfur released as SO_2 would have converted slowly to aerosols because of the low oxidation rate of SO_2 to SO_3 in the stratosphere, the duration of sunlight shielding induced by sulfate aerosols would have been controlled by the SO_2/SO_3

Table 19.1 Estimates of sulfur released from the target and projectile and NO produced by the K–Pg impact

Projectile Diameter (km)	S from the target ^a		S from the target (10 ¹⁵ mol)	NO produced by K–Pg impact ^d	Meteorite type ^b	Density (g/cm ³)	S concentration ^e (wt%)	S from the projectile (Gt)	S from the projectile (10 ¹⁵ mol)
	Density (g/cm ³)	(Gt)							
10	3.32 ^c	45–75 ^f	1.4–2.3	0.35	CV	3.42	2.2	39	1.2
					CO	3.63	2.2	42	1.3
					CR	3.27	1.9	33	1.0
15	2.49 ^g	76–127 ^h	2.4–4.0	0.59	CV	3.42	2.2	100	3.1
					CO	3.63	2.2	106	3.3
					CR	3.27	1.9	82	2.6
20	1.66 ⁱ	117–197 ^h	3.7–6.2	0.76	CV	3.42	2.2	158	4.9
					CO	3.63	2.2	167	5.2
					CR	3.27	1.9	130	4.1

^aMaximum and minimum mass represent sulfur released from target consisting of 50%:50% and 70%:30% mixture of carbonate and evaporate, respectively

^bProjectile types inferred by KYTE (1998)

^cData from Lodders and Fegley (1998)

^dValues determined from Figure 11 in Zahnle (1990)

^eProjectile porosities assumed to be 0%

^fEstimates from Pierazzo et al. (1998)

^gProjectile porosities assumed to be 25% (Pierazzo et al. 1998, 2003)

^hEstimates from Pierazzo et al. (2003)

ⁱProjectile porosities assumed to be 50% (Pierazzo et al. 1998, 2003)

ratio of the vapor plume. Ohno et al. (2004, 2014) estimated that the SO_2/SO_3 ratio in the K–Pg impact vapor cloud was significantly less than 1 ($\sim 10^{-6}$). If all sulfur was injected into the stratosphere as SO_3 as they suggested, the duration of the temperature drop of several degrees induced by sunlight shielding has been estimated to be 2 years (Pierazzo et al. 2003).

19.3 Acid Rain

Heavy acid rain immediately after the K–Pg impact explains some observations at the K–Pg boundary, such as the destruction of calcareous plankton in the oceans (Bown 2005; Jiang et al. 2010; Alegret et al. 2012), excursion of seawater $^{87}\text{Sr}/^{86}\text{Sr}$ ratios (e.g., Macdougall 1988; MacLeod et al. 2001; Frei and Frei 2002), etched pits on spinel surfaces (Preisinger et al. 2002), and enhanced sulfide accumulation in nonmarine K–Pg strata (Maruoka et al. 2002).

Sulfuric acid rain might have occurred after the K–Pg impact. Sulfur amounts released from the target and projectile (Table 19.1) correspond to $(2.4\text{--}3.7) \times 10^{15}$ or $(4.9\text{--}7.3) \times 10^{15}$ moles of sulfate for an asteroid of 10 or 15 km in diameter, respectively. The amount would have resulted in a sulfuric acid concentration 0.5–0.7 or 1.0–1.4 kg/m^2 , respectively. Even if the maximum residence time of 10 years for sulfate aerosols in the stratosphere (Pierazzo et al. 2003) is adopted, the amounts of sulfuric acid supplied to the environment would have been more than one order of magnitude higher than the “critical load” for the least sensitive freshwater environment (0.002 $\text{kg}/\text{m}^2/\text{year}$; Jeffries et al. 1999). A “critical load” is defined as the highest deposition of acidifying compounds that will not cause chemical changes leading to long-term harmful effects on the ecosystem structure and function (Nilsson and Grennfelt 1988). If the acid supply reaches a value higher than the critical load, the biota of acidified lakes might not recover completely even after recovery of the water quality (Keller et al. 1992).

In addition, acid rain from nitric acid (HNO_3) might have occurred after the K–Pg impact. Nitric oxide (NO) could have been produced by the interaction of the atmosphere with the meteorite and the ejecta (Lewis et al. 1982; Prinn and Fegley 1987; Zahnle 1990). This NO would have been converted to nitrogen dioxide (NO_2) on the time scale of minutes to hours and then to nitric acid on a time scale of about a week (Crutzen 1979). Parkos et al. (2015) estimated that the amount of NO produced during the ejecta reentry was 1.5×10^{14} moles. Such nitric acids should have not affected the acidity of the hydrosphere after the K–Pg impact because the amounts of nitric acids generated by the impact would have been one order of magnitude less than the amounts of sulfuric acids generated by the impact. On the other hand, as nitrate is one of the limiting nutrients in the ocean, the resulting nitrates could have been sufficient to facilitate eutrophication. The isotopic anomalies of nitrogen (Gilmour and Boyd 1988; Gilmour et al. 1990; Gardner et al. 1992) and carbon (Maruoka et al. 2007) induced by such eutrophication were observed for the K–Pg boundary clays. In addition, nitrogen oxide reacts with ozone in the

stratosphere, which leads to ozone destruction and, therefore, this would have increased the ultraviolet radiation. Ozone destruction in the stratosphere induced by these phenomena will be discussed in Sect. 19.5.1.

The combination of sulfuric and nitric acid rain may not have been sufficient to acidify the ocean basins (D'Hondt et al. 1994; Tyrrell et al. 2015). However, this would have caused effects on shallow and/or poorly buffered estuaries and continental catchments and waterways (Bailey et al. 2005; Ohno et al. 2014). Although the acid rain would be expected to have more seriously affected freshwater environments than marine environments (e.g., Gorham 1998), only a minor extinction of freshwater species at the K–Pg boundary is evident. Sheehan and Fastovsky (1992) inferred a 10% extinction for freshwater vertebrates, whereas an 88% extinction was found for land-dwelling vertebrates. However, such extinction selectivity can be explained by the effects of impact-generated buffers such as calcareous smectitic soils (Retallack 1996), larnite (Ca_2SiO_4) (Maruoka and Koeberl 2003), and calcite condensates from the vapor plume produced by the K–Pg impact (Schulte et al. 2009) (Fig. 19.5).

19.4 Global Warming Caused by Carbon Dioxide

19.4.1 CO_2 Release After the K–Pg Impact

Water and CO_2 were released from the target lithology at the K–Pg impact, which could have potentially caused greenhouse warming after dusts, aerosols, and soot settled to the ground (e.g., O'Keefe and Ahrens 1989; Takata and Ahrens 1994; Pierazzo et al. 1998). Estimates of CO_2 released from the target vary from 260 Gt (Ivanov et al. 1996) to 100,000 Gt (Takata and Ahrens 1994). The larger values may be an overestimate because the total released amount was most likely reduced by rapid back-reactions in which volatilized CaO and CO_2 recombined to reform CaCO_3 within the impact vapor plume (Agrinier et al. 2001). However, it is difficult to estimate how much of the CO_2 recombined to CaCO_3 in the Chicxulub impact plume. In addition, about 900 Gt of CO_2 are estimated to have been released by global wildfires after the K–Pg impact (Ivany and Salawitch 1993; Arinobu et al. 1999), although the intensity and duration of such wildfires are still poorly understood so far, as suggested in Sect. 19.2.2. Thus, the amount of CO_2 supplied to the atmosphere immediately after the K–Pg impact is still uncertain. The pre-impact atmospheric $p\text{CO}_2$ level is estimated to range from two to ten times the present atmospheric level (PAL) (Andrews et al. 1995; Berner 1998; Retallack 2001). Note that the preindustrial atmospheric level of CO_2 was 280 ppmv, which corresponds to about 2000 Gt of CO_2 . Because the estimates for the post-impact global temperature depend on which pre- and post-impact $p\text{CO}_2$ values are used, it is difficult to estimate the magnitude of global warming after the K–Pg impact.

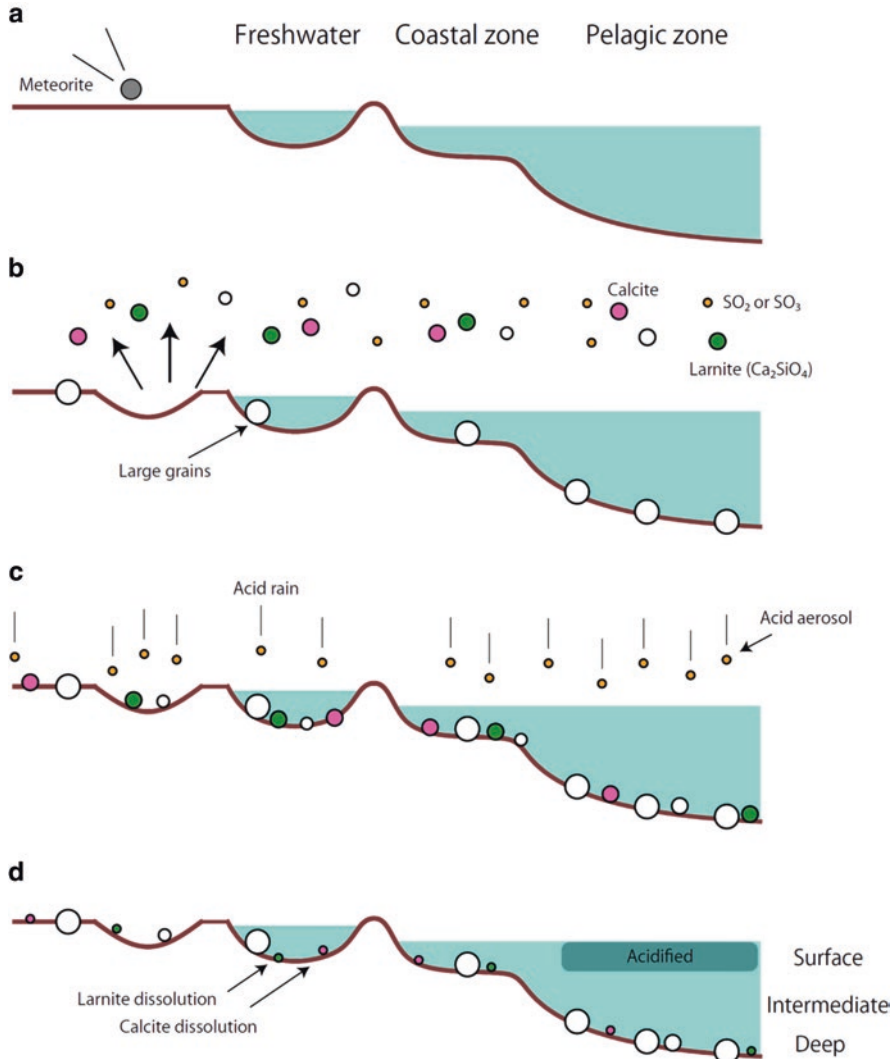


Fig. 19.5 Schematic drawing depicting the acid buffer scenario (After Maruoka and Koeberl 2003): (a) before the K-Pg impact event, (b) immediately after the impact, (c) acid aerosol deposition as acid rain, and (d) after acid neutralization. Species that rely on freshwater or nearshore ecosystems have been protected from acid rain by the acid-neutralization reaction of minerals such as carbonate and larnite distributed by the meteorite impact (d). In the ocean, solid condensates quickly settle to the bottom, whereas the acid remains near surface due to immiscibility of water with different salinities. Therefore, acid rapidly separated from neutralizer, such as calcite and larnite, at the surface layers, which led to life being more severely affected in the surface layers (d)

19.4.2 pCO₂ Estimates Before and After the K–Pg Impact

Although pre- and post-impact atmospheric pCO₂ levels have been estimated by using proxies for atmospheric pCO₂, controversial results have been obtained so far. Based on the stomatal index of land plant leaves (i.e., percentage of leaf epidermal cells that are stomata), Beerling et al. (2002) reported an extremely elevated pCO₂ level of 2300 ppmv at the K–Pg boundary, whereas the Late Cretaceous/Early Paleogene background levels were 350–500 ppmv. They suggested that there could have been a ~7.5 °C increase in the global average temperature in the absence of counter forcing by sulfate aerosols. However, as the K–Pg pCO₂ value was reconstructed by using material from a completely different plant group from the rest, it will be essential to verify that there are no systematic biases depending on plant species. Therefore, the pCO₂ spike at the K–Pg boundary proposed from the stomatal index still needs to be investigated (see Steinthorsdottir et al. 2016).

Using a pedogenic CO₂ paleobarometer, Nordt et al. (2002) suggested that there was a dramatic rise of pCO₂ from ~800 ppmv in the Maastrichtian (the latest age of the Late Cretaceous epoch) to ~1400 ppmv before the K–Pg boundary. But then, the levels are thought to have declined sharply back to ~800 ppmv at the boundary, thus suggesting that there was no CO₂ spike at the K–Pg boundary. It has been argued that the pCO₂ drawdown at the K–Pg boundary might reflect a cessation of volcanic activity in the Deccan Traps. However, because there are deficiencies in the knowledge on the key parameters used for pedogenic CO₂ paleobarometry, such as the concentrations of CO₂ supplied by soil respiration, this paleobarometry method is thought to be limited in its ability to determine ancient CO₂ levels with precision (Retallack 2009; Royer 2010; Cotton and Sheldon 2012). Thus, this result also needs to be tested (see Huang et al. 2013).

19.4.3 Temperature Variation Estimated from the Sedimentary Rock

Paleotemperature reconstructions based on carbonate δ¹⁸O values have been applied to K–Pg boundary sections. Most studies involving paleotemperature reconstruction are based on a variety of carbonate rocks and different species of planktonic and benthic foraminifera in different degrees of diagenesis (e.g., Kaiho and Lamolda 1999). It has been pointed out that isotope curves based on whole rock data from shallow water limestone with a varied diagenetic history must be carefully reconsidered (Banner and Hanson 1990; Marshall 1992). Carbonate δ¹⁸O values have been analyzed for the well-preserved K–Pg carbonate samples (Fig. 19.6; e.g., Shackleton and Hall 1984; Zachos et al. 1985). These results revealed a minor (short-term) δ¹⁸O increase across the K–Pg boundary, which is suggestive of a minor cooling event that might have been induced by sunlight shielding. However, evidence for CO₂-induced global warming (i.e., δ¹⁸O decrease) has not been found, although such

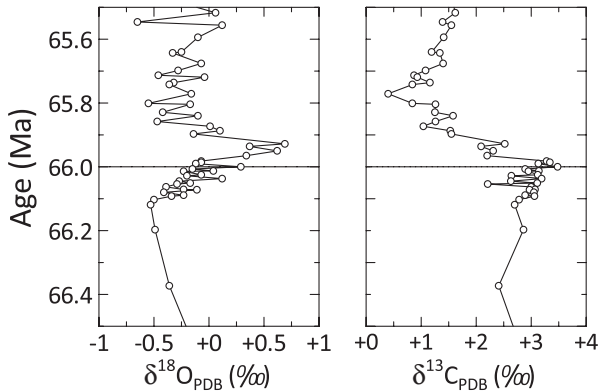


Fig. 19.6 Oxygen and carbon isotopic compositions of carbonate across the K–Pg boundary for site DSDP 527 (Data are from Shackleton and Hall 1984). $\delta^{18}\text{O}$ values for carbonate can be a temperature proxy of ocean water, whereas $\delta^{13}\text{C}$ values for carbonate can be a proxy for organic burial that is controlled by photosynthetic organic matter productivity and organic matter preservation. The $\delta^{18}\text{O}$ values increased above the K–Pg boundary, suggesting a temperature drop over a short period. The $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ values are reported relative to the PDB (Pee Dee Belemnite) standard

evidence for global warming, if it were to manifest, would be observed more easily because of its longer duration. Vellekoop et al. (2014, 2016) reported a similar trend (i.e., minor cooling and no indication of warming) using TEX86 paleothermometry. The TEX86 paleothermometry approach is based upon the SST (sea surface temperature) dependence of chemical compositions of membrane lipids produced by marine Thaumarchaeota (Schouten et al. 2002). Thus, there is no clear evidence for CO_2 -induced global warming after the K–Pg impact.

19.5 Ozone Destruction and Production

19.5.1 Ozone Destruction in the Stratosphere

The ozone layer in the stratosphere might have been destroyed by chlorine and bromine produced from the vaporized projectile, vaporized target lithology, and biomass burning (Kring et al. 1995; Kring 1999; Kourtidis 2005). Kring (1999) estimated that the amount of Cl was over five orders of magnitude more than that needed to destroy today's ozone layer, and this might have occurred along with increases in Br and other reactants. NO produced by the ejecta reentry (Lewis et al. 1982; Prinn and Fegley 1987; Zahnle 1990) also had the capacity to destroy ozone (Toon et al. 1997). However, it is difficult to evaluate how the ozone layer destruction would have affected the surface ultraviolet conditions. Ultraviolet radiation is absorbed by dust, soot, and NO_2 , and it is scattered by sulfate aerosols (Kring 1999). Thus, the ultraviolet penetration possibly depended on their settling times. The

settling time of clastic dust was probably rapid relative to the time span of ozone loss, but it may have taken a few years for the sulfate aerosols to precipitate (Pierazzo et al. 2003). Thus, the ultraviolet penetration might have been mitigated by sulfate aerosols during the first few years during the maximum depletion period for the ozone layer in the stratosphere.

19.5.2 Ozone Production in the Troposphere

Ozone occurs naturally also in the troposphere. Tropospheric ozone (ground-level ozone) acts as a powerful respiratory irritant. Enhanced production of ground-level ozone can be expected because of the high concentrations of NO_x and CH₄ supplied by the K–Pg impact (Kikuchi and Vanneste 2010). The former could have been produced by the interaction of the atmosphere with the impact projectile and ejecta, whereas the latter might have been produced by biomass burning, vaporization from the target, and destruction of methane hydrates. Kikuchi and Vanneste (2010) suggested that the ground-level O₃ concentrations reached ~1000 ppb at the maximum level after the K–Pg impact, which might have posed threats to land-dwelling life. Kawaragi et al. (2009) also suggested that there was an increase of ground-level ozone as a result of its production through photochemical reactions associated with CO. Based on experimental results, they proposed that the dissociation of CaCO₃ produced CO rather than CO₂ at the K–Pg impact.

Ozone production is driven photochemically at ground level, and its production rate would have been influenced by the penetration of ultraviolet radiation. Such penetration could have been influenced by the thickness of stratospheric ozone. Therefore, it might have been affected by concentrations of NO_x and halogens in the stratosphere and the amount of submicron-size grains such as clastic dusts, soot, and sulfate aerosols, which can absorb or scatter ultraviolet radiation. Thus, many environmental perturbations discussed in this chapter are related with each other.

19.6 Conclusion

The Mesozoic Era, which ended 66.0 million years ago, was marked by a mass extinction event that was triggered by a meteorite impact at Chicxulub on the Yucatán Peninsula in Mexico. The potential environmental perturbations following the Chicxulub impact were inferred as follows:

1. Sunlight shielding by submicron-size grains. This phenomenon was mediated by soot particles and/or sulfate aerosols. Soot may have been produced by global-scale wildfires, and sulfate aerosols were produced from sulfur-containing gases released from the impact target and the meteorite.

2. Acid rain. This was due to sulfuric acid produced from the sulfate aerosols and nitric acid produced by the interaction of the atmosphere with the meteorite and ejecta.
3. Global warming caused by the carbon dioxide released from the carbonate target of the impact.
4. Ultraviolet exposure induced by ozone destruction in the stratosphere. Ozone destruction was induced by nitric oxide produced by the interaction of the atmosphere with the meteorite and ejecta and/or by halogens (Cl and Br) released from the target materials.
5. Toxic effects of ground-level ozone. Ozone in the troposphere was produced by photochemical reactions involving hydrocarbons, nitrogen oxides, and carbon monoxides, which were supplied to the atmosphere by the meteorite impact.

Evidence for some events, such as sunlight shielding, global wildfires, and acid rain, can be found in sedimentary rocks. However, it is difficult to reconstruct quantitatively their durations and magnitudes only from the sedimentary rocks because of the deficiency in time-resolved information (yearly to millennium-scale). These events might have been influenced by each other. For instance, the sulfate aerosol production rate might have been influenced by the amount of submicron-size clastic grains and soot in the stratosphere because those grains can enhance sulfate coagulation. Numerous interactions should thus be incorporated into model calculations to reconstruct the proper environmental conditions before, during, and after the K–Pg boundary.

References

- Agrinier P, Deutsch A, Schärer U, Martinez I (2001) Fast back-reactions of shock-released CO₂ from carbonates: an experimental approach. *Geochim Cosmochim Acta* 65:2615–2632
- Alegret L, Thomas E, Lohmann KC (2012) End-Cretaceous marine mass extinction not caused by productivity collapse. *Proc Natl Acad Sci* 109:728–732
- Alvarez LW, Alvarez W, Asaro F, Michel HV (1980) Extraterrestrial cause for the Cretaceous–Tertiary extinction. *Science* 208:1095–1108
- Andrews JE, Tandon SK, Dennis PF (1995) Concentration of carbon dioxide in the Late Cretaceous atmosphere. *J Geol Soc* 152:1–3
- Arinobu T, Ishiwatari R, Kaiho K, Lamolda MA (1999) Spike of pyrosynthetic polycyclic aromatic hydrocarbons associated with an abrupt decrease in $\delta^{13}\text{C}$ of a terrestrial biomarker at the Cretaceous–Tertiary boundary at Caravaca, Spain. *Geology* 27:723–726
- Bailey JV, Cohen AS, Kring DA (2005) Lacustrine fossil preservation in acidic environments: implications of experimental and field studies for the Cretaceous–Paleogene boundary acid rain trauma. *Palaios* 20:376–389
- Banner JL, Hanson GN (1990) Calculation of simultaneous isotopic and trace-element variations during water-rock interaction with applications to carbonate diagenesis. *Geochim Cosmochim Acta* 54:3123–3137
- Beerling DJ, Lomax BH, Royer DL et al (2002) An atmospheric pCO₂ reconstruction across the Cretaceous–Tertiary boundary from leaf megafossils. *Proc Natl Acad Sci* 99:7836–7840
- Belcher CM, Collinson ME, Sweet AR et al (2003) Fireball passes and nothing burns—the role of thermal radiation in the Cretaceous–Tertiary event: evidence from the charcoal record of North America. *Geology* 31:1061–1064

- Belcher CM, Collinson ME, Scott AC (2005) Constraints on the thermal power released from the Chicxulub impactor: new evidence from multi-method charcoal analysis. *J Geol Soc Lond* 162:591–602
- Belcher CM, Finch P, Collinson ME et al (2009) Geochemical evidence for combustion of hydrocarbons during the K–T impact event. *Proc Natl Acad Sci* 106:4112–4117
- Belcher CM, Hadden RM, Rein G et al (2015) An experimental assessment of the ignition of forest fuels by the thermal pulse generated by the Cretaceous–Palaeogene impact at Chicxulub. *J Geol Soc* 172:175–185
- Berner RA (1998) The carbon cycle and carbon dioxide over Phanerozoic time: the role of land plants. *Philos Trans R Soc Lond Ser B Biol Sci* 353:75–82
- Bohor BF, Foord EE, Modreski PJ, Triplehorn DM (1984) Mineralogic evidence for an impact event at the Cretaceous–Tertiary boundary. *Science* 224:867–869
- Bohor BF, Foord EE, Ganapathy R (1986) Magnesioferrite from the Cretaceous–Tertiary boundary, Caravaca, Spain. *Earth Planet Sci Lett* 81:57–66
- Bohor BF, Modreski PJ, Foord EE (1987) Shocked quartz in the Cretaceous–Tertiary boundary clays: evidence for a global distribution. *Science* 236:705–709
- Bown PR (2005) Selective calcareous nannoplankton survivorship at the Cretaceous–Tertiary boundary. *Geology* 33:653–656
- Brett R (1992) The Cretaceous–Tertiary extinction: a lethal mechanism involving anhydrite target rocks. *Geochim Cosmochim Acta* 56:3603–3606
- Carlisle DB, Braman DR (1991) Nanometre-size diamonds in the Cretaceous/Tertiary boundary clay of Alberta. *Nature* 352:708–709
- Cotton JM, Sheldon ND (2012) New constraints on using paleosols to reconstruct atmospheric pCO₂. *Geol Soc Am Bull* 124:1411–1423
- Crocket JH, Paul DK (2004) Platinum-group elements in Deccan mafic rocks: a comparison of suites differentiated by Ir content. *Chem Geol* 208:273–291
- Crutzen PJ (1979) The role of NO and NO₂ in the chemistry of the troposphere and stratosphere. *Annu Rev Earth Planet Sci* 7:443–472
- D’Hondt S, Pilson MEQ, Sigurdsson H et al (1994) Surface-water acidification and extinction at the Cretaceous–Tertiary boundary. *Geology* 22:983–986
- Frei R, Frei KM (2002) Multi-isotopic and trace element investigation of the Cretaceous–Tertiary boundary layer at Stevns Klint, Denmark – inferences for the origin and nature of siderophile and lithophile element geochemical anomalies. *Earth Planet Sci Lett* 203:691–708
- Ganapathy R (1980) A major meteorite impact on the earth 65 million years ago: evidence from the Cretaceous–Tertiary boundary clay. *Science* 209:921–923
- Gardner A, Hildebrand A, Gilmour I (1992) Isotopic composition and organic geochemistry of nitrogen at the Cretaceous/Tertiary boundary. *Meteoritics* 27:222–223
- Gilmour I, Boyd S (1988) Nitrogen geochemistry of a Cretaceous–Tertiary boundary site in New Zealand. *LPI Contrib* 673:58–59
- Gilmour I, Wolbach WS, Anders E (1990) Early environmental effects of the terminal Cretaceous impact. *Geol Soc Am Spec Pap* 247:383–390
- Gilmour I, Russel SS, Arden JW et al (1992) Terrestrial carbon and nitrogen isotopic-ratios from Paleogene–Tertiary boundary nanodiamonds. *Science* 258:1624–1626
- Gorham E (1998) Acid deposition and its ecological effects: a brief history of research. *Environ Sci Pol* 1:153–166
- Hildebrand AR, Penfield GT, Kring DA, Boynton WV (1991) Chicxulub crater: a possible Cretaceous/Tertiary boundary impact crater on the Yucatan Peninsula, Mexico. *Geology* 19:867–871
- Huang C, Retallack GJ, Wang C, Huang Q (2013) Paleatmospheric pCO₂ fluctuations across the Cretaceous–Tertiary boundary recorded from paleosol carbonates in NE China. *Palaeogeogr Palaeoclimatol Palaeoecol* 385:95–105
- Ivanov BA, Badukov DD, Yakovlev OI et al (1996) Degassing of sedimentary rocks due to Chicxulub impact: hydrocode and physical simulations. *Geol Soc Am Spec Pap* 307:125–139
- Ivany LC, Salawitch RJ (1993) Carbon isotopic evidence for biomass burning at the K–T boundary. *Geology* 21:487–490

- Izett GA (1991) Tektites in Cretaceous-Tertiary boundary rocks on Haiti and their bearing on the Alvarez Impact Extinction Hypothesis. *J Geophys Res Planet* 96:20879–20905
- Jeffries DS, Lam DCL, Moran MD, Wong I (1999) Effect of SO₂ emission controls on critical load exceedances for lakes in southern Canada. *Water Sci Technol* 39:165–171
- Jiang S, Bralower TJ, Patzkowsky ME et al (2010) Geographic controls on nanoplankton extinction across the Cretaceous/Paleogene boundary. *Nat Geosci* 3:280–285
- Jones TP, Lim B (2000) Extraterrestrial impacts and wildfires. *Palaeogeogr Palaeoclimatol Palaeoecol* 164:57–66
- Kaiho K, Lamolda MA (1999) Catastrophic extinction of planktonic foraminifera at the Cretaceous–Tertiary boundary evidenced by stable isotopes and foraminiferal abundance at Caravaca, Spain. *Geology* 27:355–358
- Kaiho K, Oshima N, Adachi K et al (2017) Global climate change driven by soot at the K–Pg boundary as the cause of the mass extinction. *Sci Rep* 6:28427
- Kawaragi K, Sekine Y, Kadono T et al (2009) Direct measurements of chemical composition of shock-induced gases from calcite: an intense global warming after the Chicxulub impact due to the indirect greenhouse effect of carbon monoxide. *Earth Planet Sci Lett* 282:56–64
- Keller W, Gunn JM, Yan ND (1992) Evidence of biological recovery in acid stressed lakes near Sudbury, Canada. *Environ Pollut* 78:79–85
- Kikuchi R, Vanneste M (2010) A theoretical exercise in the modeling of ground-level ozone resulting from the K–T asteroid impact: its possible link with the extinction selectivity of terrestrial vertebrates. *Palaeogeogr Palaeoclimatol Palaeoecol* 288:14–23
- Kourtidis K (2005) Transfer of organic Br and Cl from the biosphere to the atmosphere during the Cretaceous/Tertiary impact: implications for the stratospheric ozone layer. *Atmos Chem Phys* 5:207–214
- Kring DA (1999) Ozone-depleting chlorine and bromine produced by the Chicxulub impact event. *Meteorit Planet Sci* 34:A67–A68
- Kring DA, Melosh HJ, Hunten DM (1995) Possible climatic perturbations produced by impacting asteroids and comets. *Meteoritics* 30:530
- Kyte FT (1998) A meteorite from the Cretaceous/Tertiary boundary. *Nature* 396:237–239
- Kyte FT, Smit J (1986) Regional variations in spinel compositions: an important key to the Cretaceous/Tertiary event. *Geology* 14:485–487
- Lewis J, Watkins GH, Hartman H, Prinn R (1982) Chemical consequences of major impact events on Earth. *Geol Soc Am Spec Pap* 190:215–221
- Lodders K, Fegley B Jr (1998) *The planetary scientist's companion*. Oxford University Press, New York, p 371
- López-Ramos E (1975) Geological summary of Yucatán Peninsula. In: Nairn AEM, Stehli FG (eds) *The ocean basins and margins, vol 3 The Gulf of Mexico and the Caribbean*. Plenum Press, New York, pp 257–282
- Macdougall JD (1988) Seawater strontium isotopes, acid rain, and the Cretaceous–Tertiary boundary. *Science* 239:485–487
- MacLeod KG, Huber BT, Fullagar PD (2001) Evidence for a small (~0.000030) but resolvable increase in seawater ⁸⁷Sr/⁸⁶Sr ratios across the Cretaceous–Tertiary boundary. *Geology* 29:303–306
- Marshall JD (1992) Climatic and oceanographic isotopic signals from the carbonate rock record and their preservation. *Geol Mag* 129:143–160
- Maruoka T, Koeberl C (2003) Acid-neutralizing scenario after the K–T impact event. *Geology* 31:489–492
- Maruoka T, Koeberl C, Newton J et al (2002) Sulfur isotopic compositions across terrestrial Cretaceous–Tertiary (K–T) boundary successions. *Geol Soc Am Spec Pap* 356:337–344
- Maruoka T, Koeberl C, Bohor BF (2007) Carbon isotopic compositions of organic matter across continental Cretaceous–Tertiary (K–T) boundary sections: implications for paleoenvironment after the K–T impact event. *Earth Planet Sci Lett* 253:226–238
- McHone JF, Nieman RA, Lewis CF, Yates AM (1989) Stishovite at the Cretaceous–Tertiary boundary, Raton, New Mexico. *Science* 243:1182–1184

- Melosh HJ, Schneider NM, Zahnle K, Latham D (1990) Ignition of global wildfires at the Cretaceous/Tertiary boundary. *Nature* 343:251–254
- Morgan J, Artemieva N, Goldin T (2013) Revisiting wildfires at the K–Pg boundary. *J Geophys Res Biogeosci* 118:1508–1520
- Nilsson J, Grennfelt P (1988) Critical loads for sulfur and nitrogen. Nordic Council of Ministers, Copenhagen
- Nordt L, Atchley S, Dworkin SI (2002) Paleosol barometer indicates extreme fluctuations in atmospheric CO₂ across the Cretaceous–Tertiary boundary. *Geology* 30:703–706
- O’Keefe JD, Ahrens TJ (1989) Impact production of CO₂ by Cretaceous/Tertiary extinction bolide and the resultant heating of the Earth. *Nature* 338:247–249
- Ohno S, Sugita S, Kadono T, Hasegawa S, Igarashi G (2004) Sulfur chemistry in laser-simulated impact vapor clouds: implications for the K/T impact event. *Earth Planet Sci Lett* 218:347–361
- Ohno S, Kadono T, Kurosawa K et al (2014) Production of sulphate-rich vapour during the Chicxulub impact and implications for ocean acidification. *Nat Geosci* 7:279–282
- Palme H, Jones A (2003) Solar system abundances of the elements. In: Davis AM (ed) *Treatise on geochemistry 1. Meteorites, comets and planets*. Elsevier, Amsterdam, pp 41–61
- Parkos D, Alexeenko A, Kulakhmetov M et al (2015) NO_x production and rainout from Chicxulub impact ejecta and reentry. *J Geophys Res Planet* 120:2152–2168
- Pierazzo E, Kring DA, Melosh HJ (1998) Hydrocode simulation of the Chicxulub impact event and the production of climatically active gases. *J Geophys Res* 103:28,607–28,625
- Pierazzo E, Hahmann AN, Sloan LC (2003) Chicxulub and climate: effects of stratospheric injections of impact-produced S-bearing gases. *Astrobiology* 3:99–118
- Pope KO (2002) Impact dust not the cause of the Cretaceous–Tertiary mass extinction. *Geology* 30:99–102
- Preisinger A, Aslanian S, Brandstätter F et al (2002) Cretaceous–Tertiary profile, rhythmic deposition, and geomagnetic polarity reversal of marine sediments near Bjala, Bulgaria. *Geol Soc Am Spec Pap* 356:213–229
- Prinn RG, Fegley B Jr (1987) Bolide impacts, acid rain, and biospheric traumas at the Cretaceous–Tertiary boundary. *Earth Planet Sci Lett* 83:1–15
- Retallack GJ (1996) Acid trauma at the Cretaceous–Tertiary boundary in eastern Montana. *GSA Today* 6:1–7
- Retallack GJ (2001) A 300 million year record of atmospheric carbon dioxide from fossil plant cuticles. *Nature* 411:287–290
- Retallack GJ (2009) Refining a pedogenic–carbonate CO₂ paleobarometer to quantify a middle Miocene greenhouse spike. *Palaeogeogr Palaeoclimatol Palaeoecol* 281:57–65
- Robertson DS, McKenna MC, Toon OB et al (2004) Comment on fireball passes and nothing burns—the role of thermal radiation in the Cretaceous–Tertiary event: evidence from the charcoal record of North America. *Geology* 32:e50
- Robertson DS, Lewis WM, Sheehan PM, Toon OB (2013) K–Pg extinction: reevaluation of the heat-fire hypothesis. *J Geophys Res Biogeosci* 118:329–336
- Robinson N, Ravizza G, Coccioni R et al (2009) A high-resolution marine ¹⁸⁷Os/¹⁸⁸Os record for the late Maastrichtian: distinguishing the chemical fingerprints of Deccan volcanism and the KP impact event. *Earth Planet Sci Lett* 281:159–168
- Royer DL (2010) Fossil soils constrain ancient climate sensitivity. *Proc Natl Acad Sci U S A* 107:517–518
- Rudnick RL, Gao S (2003) Composition of the continental crust. In: Rudnick RL (ed) *Treatise on Geochemistry 3. The crust*. Elsevier, Amsterdam, pp 1–64
- Schoene B, Samperton KM, Eddy MP et al (2015) U–Pb geochronology of the Deccan Traps and relation to the end-Cretaceous mass extinction. *Science* 347:182–184
- Schouten S, Hopmans EC, Schefuss E et al (2002) Distributional variations in marine crenarchaeotal membrane lipids: a new organic proxy for reconstructing ancient sea water temperatures? *Earth Planet Sci Lett* 204:265–274

- Schulte P, Deutsch A, Salge T et al (2009) A dual-layer Chicxulub ejecta sequence with shocked carbonates from the Cretaceous–Paleogene (K–Pg) boundary, Demerara Rise, western Atlantic. *Geochim Cosmochim Acta* 73:1180–1204
- Schulte P, Alegret L, Arenillas I et al (2010) The Chicxulub asteroid impact and mass extinction at the Cretaceous–Paleogene boundary. *Science* 327:1214–1218
- Sepkoski JJ Jr (1996) Patterns of Phanerozoic extinction: a perspective from global data bases. In: Walliser OH (ed) *Global events and event stratigraphy*. Springer, Berlin, pp 35–51
- Shackleton NJ, Hall MA (1984) Carbon isotope data from Leg 74 sediments. *DSDP Init Rep* 74:613–619
- Sheehan PM, Fastovsky DE (1992) Major extinctions of land-dwelling vertebrates at the Cretaceous–Tertiary boundary, eastern Montana. *Geology* 20:556–560
- Sigurdsson H, D’Hondt S, Arthur MA et al (1991) Glass from the Cretaceous/Tertiary boundary in Haiti. *Nature* 349:482–487
- Sigurdsson H, D’Hondt S, Carey S (1992) The impact of the Cretaceous/Tertiary bolide on evaporite terrane and generation of major sulfuric acid aerosol. *Earth Planet Sci Lett* 109:543–559
- Stanley SM (2016) Estimates of the magnitudes of major marine mass extinctions in earth history. *Proc Natl Acad Sci U S A* 113:E6325–E6334
- Steinthorsdottir M, Vajda V, Pole M (2016) Global trends of pCO₂ across the Cretaceous–Paleogene boundary supported by the first Southern Hemisphere stomatal proxy-based pCO₂ reconstruction. *Palaeogeogr Palaeoclimatol Palaeoecol* 464:143–152
- Stinnesbeck W, Keller G, Adatte T et al (2004) Yaxcopoil-1 and the Chicxulub impact. *Int J Earth Sci* 93:1042–1065
- Swisher CC III, Grajales-Nishimura JM, Montanari A et al (1992) Coeval ⁴⁰Ar/³⁹Ar ages of 65.0 million years ago from Chicxulub crater melt rock and Cretaceous–Tertiary boundary tektites. *Science* 257:954–958
- Takata T, Ahrens TJ (1994) Numerical simulation of impact cratering at Chicxulub and the possible causes of KT catastrophe. *Lunar Planet Inst Contr* 825:125–126
- Toon OB, Pollack JB, Ackerman TP et al (1982) Evolution of an impact-generated dust cloud and its effects on the atmosphere. *Geol Soc Am Spec Pap* 190:187–200
- Toon OB, Zahnle K, Morrison D et al (1997) Environmental perturbations caused by the impacts of asteroids and comets. *Rev Geophys* 35:41–78
- Tyrrell T, Merico A, McKay A, Ian D (2015) Severity of ocean acidification following the end-Cretaceous asteroid impact. *Proc Natl Acad Sci U S A* 112:6556–6561
- Urrutia-Fucugauchi J, Camargo-Zanoguera A, Pérez-Cruz L, Pérez-Cruz G (2011) The Chicxulub multi-ring impact crater, Yucatan carbonate platform, Gulf of Mexico. *Geofis Int* 50:99–127
- Vellekoop J, Sluijs A, Smit J et al (2014) Rapid short-term cooling following the Chicxulub impact at the Cretaceous–Paleogene boundary. *Proc Natl Acad Sci U S A* 111:7537–7541
- Vellekoop J, Esmeray-Senlet S, Miller KG et al (2016) Evidence for Cretaceous–Paleogene boundary bolide “impact winter” conditions from New Jersey, USA. *Geology* 44:619–622
- Ward WC, Keller G, Stinnesbeck W, Adatte T (1995) Yucatán subsurface stratigraphy: implications and constraints for the Chicxulub impact. *Geology* 23:873–876
- Wolbach WS, Lewis RS, Anders E (1985) Cretaceous extinctions: evidence for wildfires and search for meteoritic material. *Science* 230:167–170
- Wolbach WS, Gilmour I, Anders E (1990) Major wildfires at the Cretaceous/Tertiary boundary. *Geol Soc Am Spec Pap* 247:391–400
- Zachos JC, Arthur MA, Thunell RC et al (1985) Stable isotope and trace element geochemistry of carbonate sediments across the Cretaceous/Tertiary boundary at Deep Sea Drilling Project Hole 577, Leg 86. *DSDP Init Rep* 86:513–532
- Zahnle KJ (1990) Atmospheric chemistry by large impacts. *Geol Soc Am Spec Pap* 247:271–288

Part V
Search for Life in Solar System and Extra
Solar System

Chapter 20

Limits of Terrestrial Life and Biosphere



Ken Takai

Abstracts The origin, evolution, and distribution of life in the Universe can be better addressed by understanding the limits of life on Earth. A broad range of physical and chemical constraints for limits of life, such as temperature, pressure, pH, salinity, physical space, water content, and availability of energy and nutrients, have been explored in many environments of the Earth that potentially host certain boundaries between the habitable and the uninhabitable terrains. In this chapter, the presently known limits of life are described, and the possible environments and their physical and chemical characteristics that could signify the limits of life and biosphere on the Earth are reviewed. Although the nature and distribution of fringe biospheres that face the boundaries are highly unknown, numerous geomicrobiological explorations have demonstrated the limits of biosphere realistically occur in the subsurface environments and are controlled by certain boundary conditions. Energetic aspects of boundary conditions are quite important and are discussed based on the theoretical estimations and field observations of earthly life and biosphere.

Keywords Limits of life · Limits of biosphere · Fringe biosphere · Habitability · Energetic balance of habitability

20.1 Introduction

Although the concept of habitability has many aspects, it is widely accepted that the most extreme conditions of habitats, the so-called fringe biospheres, exist in deep subsurface environments (Takai 2011; Takai et al. 2014). Potential physical and chemical constraints dictating the limits of life are considered to include temperature, pressure, pH, salinity, physical space, liquid water, sufficient access to nutrients and energy sources, and so on. Adapting to these constraints, numerous

K. Takai (✉)

Department of Subsurface Geobiological Analysis and Research (D-SUGAR), Japan Agency for Marine-Earth Science & Technology (JAMSTEC), Yokosuka, Japan
e-mail: kent@jamstec.go.jp

extremophilic microbial communities have been found to occur near the boundary between the habitable and uninhabitable terrains in the deep and dark niches on Earth. These microbial communities in the deep subsurface fringe biospheres may have specific abilities to adapt and/or survive under the environmental conditions including such constraints. Thus, exploration of the fringe biospheres and elucidation of the boundary conditions provide clues to determining the conditions that may also have allowed the birth and propagation of primordial life in the early Earth and the possible existence of extraterrestrial life beyond this planet.

In this chapter, we first describe the physical and chemical characteristics of fringe biospheres in the Earth and of the previously known limits to earthly life based on the laboratory-based microbial growth experiments. Although many of the pure-culture microbiological experiments have only addressed one particular environmental constraint at a time, there are often multiple physical, chemical, and biological factors simultaneously and dynamically shaping the habitability in natural environments. In particular, microbial symbioses and synergetic functions are difficult to capture in controlled laboratory experiments. Thus, several explorations have sought to investigate the microbial communities and their functions in the in situ deep subsurface fringe biospheres. Although the nature, distribution, and function of fringe biospheres remain to be fully resolved, the geomicrobiological investigations have indicated that the indigenous microbial communities live near the limits (e.g., temperature, pH, and energy) of life and biosphere. Using the yet-limited data from such explorations, we summarize the known facts about the biotic fringes in the natural environments and discuss the differences between the boundary conditions predicted through the laboratory experiments and those observed in the natural environments.

20.2 Possible Physical and Chemical Constraints on Earthly Life

20.2.1 Temperature

In deep subsurface environments, many physical and chemical parameters can restrict the growth and sustenance of a microbial community. One of the most studied is temperature. In environments present on Earth's surface, liquid water boils at and below 100 °C, but with increasing pressure that accompanies increasing depth, liquid water can exist up to 373 °C for pure water and 407 °C for seawater (critical points) (Bischoff and Rosenbauer 1988). Indeed, the highest temperature record of liquid water (407 °C) was found in a deep-sea hydrothermal vent of the Mid-Atlantic Ridge (Table 20.1) (Koschinsky et al. 2008). Since the first discovery of high-temperature hydrothermal fluid vents in 1979 at 21°N on the East Pacific Rise (Spiess and RISE Group 1980), microbiologists have been interested in empirically determining the upper temperature limit (UTL) for life with the deep-sea vent

Table 20.1 Physical and chemical conditions of possible fringe biosphere in the Earth

Physical and chemical factor	The most extreme example in the Earth	The most extreme example in the subsurface environments	The most extreme biosphere inferred or justified
Lowest temperature	-89 °C Vostok Station in Antarctica	From -89 to 0 °C ice sheet, subglacial lake, permafrost and deep seawater	Down to -89 °C accretion ice in Lake Vostok (D’Elia et al. 2008)
	-157 °C ISS orbit in thermosphere		
Highest temperature	407 °C	407 °C	365 °C
	A deep-sea hydrothermal vent, Mid-Atlantic Ridge (Koschinsky et al. 2008)	A deep-sea hydrothermal vent, Mid-Atlantic Ridge (Koschinsky et al. 2008)	Kairei hydrothermal field, Central Indian Ridge (Takai et al. 2008a)
Lowest pressure	10 ⁻⁷ Pa	>100 KPa	0.6 KPa
	Pa ISS orbit in thermosphere		Stratosphere at 58 km altitude (Imshenetsky et al. 1976)
Highest pressure	110 MPa	110 MPa	110 MPa
	Challenger Deep, Mariana Trench (Kato et al. 1997)	Challenger Deep, Mariana Trench (Kato et al. 1997)	Challenger Deep, Mariana Trench (Kato et al. 1997)
Lowest pH	pH -3.6	pH -3.6	>pH -3.6
	A pool water of Richmond Mine, Iron Mountain (Nordstrom et al. 2000)	A pool water of Richmond Mine, Iron Mountain (Nordstrom et al. 2000)	Pool waters of Richmond Mine, Iron Mountain (Méndez-García et al. 2015)
Highest pH	pH 12.9 measured an underground water in Maqarin bitumen marl, Jordan (Pedersen et al. 2004)	pH 12.9 measured an underground water in Maqarin bitumen marl, Jordan (Pedersen et al. 2004)	>pH 12.9 an underground water in Maqarin bitumen marl, Jordan (Pedersen et al. 2004)
	pH 13.1 in situ pore waters in serpentine seamounts, Mariana Forearc (Mottl 2009)	pH 13.1 in situ pore waters in serpentine seamounts, Mariana Forearc (Mottl 2009)	
Highest salinity	Many saturated soda lakes and crystal salts	Many deep-sea saturated brine pools and crystal salts	Many saturated soda lakes, deep-sea saturated brine pools and crystal salts (Rainy and Oren 2006)
UV radiation	17.5 W/m ² for UV-B	None	Stratosphere at low-latitude area
	6.4 W/m ² for UV-C		
	ISS orbit in thermosphere (Schuster et al. 2012)		
Ionizing radiation	>1 KGy/h	>1 KGy/h	Sealed food after gamma-ray irradiation (30 KGy) (Battista 1997)
	At natural fissure reactors in the ancient Earth ~2 Ga (Dartnell 2011)	At natural fissure reactors in the ancient Earth ~2 Ga (Dartnell 2011)	

Table 20.2 Presently known limits of microbial growth and survival

Physical and chemical factor	Limit for growth and maintenance	Limit for survival
Lowest temperature	<0 °C	<0 °C for very long time
	Many psychrophiles (Rainy and Oren 2006)	Many <i>Bacteria</i> and <i>Archaea</i>
Highest temperature	122 °C at 20–40 MPa	130 °C for 180 min at 30 MPa
	<i>Methanopyrus kandleri</i> strain 116 (Takai et al. 2008a)	<i>Methanopyrus kandleri</i> strain 116 (Takai et al. 2008a)
Highest pressure	140 MPa at 6 °C	1.4 GPa at 120 °C
	<i>Colwellia marinimaniae</i> (Kusube et al. 2017)	<i>Clostridium botulinum</i> spore (Margosch et al. 2006)
Lowest pH	pH 0	<pH 0
	<i>Picrophilus oshimae</i> and <i>P. torridus</i> (Schleper et al. 1995)	
Highest pH	pH 12.4	Probably up to pH 14
	<i>Alkaliphilus transvaalensis</i> (Takai et al. 2001b)	<i>Alkaliphilus transvaalensis</i> spore
Highest salinity	Saturated	In crystal for 250 million years
	Many extreme halophiles (Rainy and Oren 2006)	<i>Bacillus sphaericus</i> (Vreeland et al. 2000)
UV radiation	Uncertain	10 ⁻⁵ survival with 2000 J/m ² of UV-C irradiation
		<i>Deinococcus radiodurans</i> (Battista 1997)
Ionizing radiation	Uncertain	10 ⁻⁵ survival with 20 K Gy of gamma-ray-dose
		<i>Deinococcus radiodurans</i> (Battista 1997)

microorganisms. At present, the highest temperature at which an organism is able to grow is 122 °C under 20 and 40 MPa of hydrostatic pressure conditions, (*Methanopyrus kandleri* strain 116; see Table 20.2) (Takai et al. 2008a). This UTL record for life shows a realistic physical constraint limiting microbial growth with the abundant energy sources under the laboratory conditions, but it may be a transient one. If new methods and techniques are developed for cultivation, growth, and detection of currently unknown hyperthermophiles, the UTL record may be extended further. On a related note, the highest survival temperature and duration of the same organism in a laboratory are several hours at around 130 °C (Table 20.2) (Takai et al. 2008a). However, the molecular signals of microorganisms and even living hyperthermophiles have been retrieved from samples of high-temperature (>250 °C) hydrothermal fluids or materials (e.g., Takai et al. 2004, 2008a, b). These molecular signals and living microorganisms are likely derived from a tiny fraction of the microbial populations that originally thrived at much lower temperatures but have survived with exposure to the higher temperatures of hydrothermal fluids for relatively short time periods (Takai et al. 2004, 2008b). To our knowledge, however,

no experiments have been undertaken yet that simulate microbial survival under natural habitat conditions: for example, using the microbial communities in the natural hydrothermal fluids and mineral deposits, with the abundant energy sources at temperatures exceeding 130 °C and at in situ hydrostatic pressures.

Although hydrothermal vents are observed at the seafloor, the networks of fissures, cracks, and permeable rocky and sedimentary layers supporting the hydrothermal fluid flows can be extensive beneath the seafloor (e.g., Wilcock and Fisher 2004). Thus, the subseafloor environments that are widespread around mid-ocean ridges (MOR), arc-back-arc (ABA) volcanoes and spreading centers, hot-spot volcanoes, and their flank regions can host enormous spaces that occupy liquid water at high-temperature (up to 407 °C) and even higher-temperature supercritical fluid (Shock 1992) (Table 20.1). Despite the relatively smaller spatial and temporal extents, the subsurface environments of terrestrial hydrothermal fields associated with volcanic activities have similar hydrogeological regimes. Most of the non-hydrothermal subsurface environments that are characterized by much lower temperatures are not at all fatal to most of the microorganisms. However, even in the non-hydrothermal environments, increasing depth elevates temperature according to indigenous geothermal gradient. Particularly in non-hydrothermal environments with relatively high geothermal gradients, the temperature range of subsurface environments at several km depths is expected to be up to 100–200 °C (Heuer et al. 2017). So far, the UTL records for life have been found in microorganisms living in terrestrial and deep-sea hydrothermal environments. In future UTL may be obtained from microorganisms living in such energy-rich, non-hydrothermal deep and hot biospheres.

20.2.2 Pressure

Elevated hydrostatic and lithostatic pressure is common in the subsurface environments and has been considered as another important physical constraint. The deepest habitat ever explored in the Earth is the Challenger Deep in the Mariana Trench, Pacific Ocean, at a water depth of ~10,900 m (Kato et al. 1997; Glud et al. 2013; Nunoura et al. 2015), which corresponds to 110 MPa of hydrostatic pressure (Table 20.1). The laboratory-based upper pressure limit (UPL) for microbial growth was known to be 130 MPa for a deep-sea psychrophilic heterotroph (strain MT41) isolated from the Challenger Deep (Yayanos 1986) and for a new deep-sea hyperthermophilic heterotroph *Thermococcus profundus* isolated from the Beebe hydrothermal field in the Mid Cayman Rise, Caribbean (Table 20.2) (Dalmaso et al. 2016). Very recently, however, the isolation of an obligate piezophilic heterotroph *Colwellia marinimaniae* from the Challenger Deep has led to the rise of the UPL to 140 MPa (Kusube et al. 2017). In addition, a great phylogenetic (based on rRNA gene sequences) and physiological diversity of microbial communities and highly active microbial metabolic functions have also been identified from the Challenger Deep sediments and waters of the Mariana Trench (Kato et al. 1997; Takami et al. 1997; Takai et al. 1999; Glud et al. 2013; Nunoura et al. 2015). A diverse

macrofauna has also been found to thrive there (Belliaev and Brueggeman 1989; Kobayashi et al. 2012), and some of them seem to be highly affected by terrigenous material input and anthropogenic pollutants even at the deepest part of the Earth far from the human activities (Jamieson et al. 2017). Thus, it is evident that even the deepest parts and the greatest pressure conditions ever explored are not fringe for the microbial and faunal community developments (Table 20.1).

Pressure (hydrostatic and lithostatic) also has an impact on the survival for life. The pressure and temperature effects on the spore survival have been long investigated in the field of food microbiology. As far as known, some bacterial spores (*Clostridium* and *Bacillus* spp.) can revive after pressure treatment at >1 GPa at 25 °C and at 1.4 GPa at 120 °C for a few minutes (Table 20.2) (e.g., Margosch et al. 2006). Based on the kinetic modeling of pressure and temperature effects on the spore viability (Margosch et al. 2006), a possible upper pressure limit for spore survival for a few minutes is expected to be below 2 GPa. This pressure corresponds to >150 km of water depth in the Earth's ocean or >50 km of lithospheric depth in the Earth's crust. On the other hand, the pressure effect on cellular survival has been investigated using very limited number of non-piezophilic bacterial strains such as *E. coli*, *Listeria monocytogenes*, and *Lactobacillus* spp. At 4 °C and 700 MPa, most of these bacterial members (if more than 10⁹ cells are processed under a pressure) are not completely dead for more than 5 min (Klotz et al. 2007). Since no similar study has been applied to facultative and obligate piezophiles, it is still uncertain how a great hydrostatic pressure of >700 MPa affects the cellular survival of extreme piezophiles dominating in the deep ocean and deep subsurface environments. However, if the UPL for cellular survival for a few minutes is possibly similar to that for spore survival, then pressure alone is not a decisive factor in controlling the survival limitation of life and biosphere in this planet.

20.2.3 Salinity and pH

Extreme salinity and pH conditions are also found in the subsurface environments. Volcanic activity often leads to occurrence of extremely acidic hot water springs on land, of which pH values can drop close to and even below 0 (e.g., Rowe et al. 1992; Schleper et al. 1995; Darrach et al. 2013). This extreme acidity is caused by inputs of sulfuric acid, hydrochloric acid, and/or hydrofluoric acid originally provided from magmatic volatiles associated with volcanic activity (Rowe et al. 1992). The acidic pH limit for growth and survival is well established for extremely acidophilic *Archaea* belonging to the phylum *Thermoplasmata* such as *Picrophilus* and *Ferroplasma* members that can grow even at pH 0 (Table 20.2) (Schleper et al. 1995; Edwards et al. 2000), and several eukaryotic species can also grow and survive at pH 0 (Doemel and Brock 1971; Rothschild and Mancinelli 2001). In the deep-sea environments, the lowest pH value ever reported is 1.6 from the TOTO caldera field in the Mariana Arc (Nakagawa et al. 2006), and more acidic values (pH <1) are observed in western Pacific submarine volcanoes (Table 20.1) (Resing et al.

2007; Butterfield et al. 2011). In addition, the lowest pH value of liquid water (pH -3.6) is ever obtained from an underground acid mine pool water in the Richmond Mine at Iron Mountain, California (Table 20.1) (Nordstrom et al. 2000). Indeed, it has been shown by metagenomic investigations that a diversity of microbial community dominated by *Leptospirillum* bacterial, *Ferroplasma* archaeal, and possible acidophilic fungal phylotypes thrive in the extremely acidic mine water environments (Méndez-García et al. 2015). The lowest pH condition alone may be not a boundary condition to determine the limits of life and biosphere in this planet. Although it is predicted that the acidic habitats for seafloor microbial communities are widespread beneath the hydrothermally active seafloor in such submarine volcanoes, the most acidophilic microorganism from the deep-sea and seafloor hydrothermal environments, '*Aciduliprofundum boonei*', grows at down to pH 3.3 (Reysenbach et al. 2006). It seems unlikely that the quite low pH values (<0) can be found in subsurface high-temperature fluid regimes due to the solubility of acids under the hydrothermal conditions (Seyfried et al. 1991).

On the other end of the pH spectrum, extremely alkaline environments are also generated in the surface and deep subsurface hydrothermal (water-rock) processes. The serpentinization reaction of water and mafic minerals, which are common in rocks present in continental ophiolite zones and in oceanic crusts along ultraslow to intermediate spreading centers, is well known for generating highly alkaline fluids and their water regimes (McCollom and Bach 2009; Schrenk et al. 2013). Although the pH values of serpentinization-driven fluids found in continental ophiolite zones usually perch on up to around pH 12 (Schrenk et al. 2013), the serpentinite mud pore waters of the serpentinite seamounts in the Mariana Forearc have the measured pH values at 25 °C as high as 12.6 (Table 20.1) (Salisbury et al. 2002; Mottl et al. 2003). This is close to the highest pH value ever measured (pH 12.9) in the extremely alkaline underground water in the Maqarin "bituminous marl formation" in Jordan (Table 20.1) (Pedersen et al. 2004), while the in situ pH values of serpentinite mud pore waters are estimated to be pH 13.1 at 2–3 °C, potentially representing the most alkaline environment in this planet (Table 20.1) (Mottl 2009). In addition, highly alkaline hydrothermal systems driven by serpentinization could potentially be widespread in off-axis environments far from magmatic activity, such as the Lost City hydrothermal field (Kelley et al. 2001, 2005), and would also provide the possible extreme habitats characterized by both high-temperature and high-pH conditions.

Alkaliphilic bacteria are known to be ubiquitous in non-alkaline habitats such as soil, freshwater, and ocean environments and are usually able to grow up to pH 10–11 (Horikoshi 1999). This pH range is almost equivalent to the alkaline pH limit for life (pH 10–12) that has been often described in the literature (Rothschild and Mancinelli 2001; Rainy and Oren 2006). However, the most alkaliphilic microorganism *Alkaliphilus transvaalensis* was isolated from an ultra-deep South African gold mine (3.2 km deep below land surface) and is known to grow at up to pH 12.4 (Table 20.2) (Takai et al. 2001b). Since *A. transvaalensis* is a gram-positive, spore-forming bacterium, the cells and spores would be able to survive under more alkaline pH conditions than the alkaline pH limit for growth. In addition, considerable biomass, cultivability, phylogenetic diversity, and anabolic function of alkaliphilic

bacterial populations were identified in the groundwater samples where the highest pH values are measured (pH 12.7–12.9) (Pedersen et al. 2004) and the bacterial growth was verified even under the alkaline pH condition of 12.5. These observations suggest that the alkaline pH limit for life may occur at pH >12.9 in certain natural subsurface environments with the abundant energy sources. However, in the subseafloor serpentinite mud environments of several Mariana Forearc serpentine seamounts, the pH values of pore waters are measured to be as high as pH 12.6 but to be pH 13.1 in situ (Table 20.1) (Salisbury et al. 2002; Mottl 2009). In the pristine, potentially the most alkaline, serpentine fluid regimes (at in situ pH 13), the lipid biomarkers and sulfur isotopic signatures have pointed at the possible occurrence of subseafloor microbial community (Mottl et al. 2003; Aoyama et al. 2018), while no living microorganism has been cultivated under any of pH conditions, and the reliable microbial biomass and anabolic activity have been scarcely detected (Takai et al. 2005; Kawagucci et al. 2018). Based on these results, it is also interpreted that the alkaline pH limit for life would occur at around pH 13 in the subsurface environments. Although any of the clear evidences of the boundary is not yet shown, the alkaline condition around in situ pH 13 is likely to be a possible alkaline limit for life and biosphere even with the abundant energy sources. Further microbiological investigations would extend the present alkaline pH limit for life and find examples of alkaline pH limit for biosphere in the subsurface environments.

Not only the surface but also the subsurface environments can host microbial habitats with a variety of salinities from almost fresh to hypersaline water and potentially even in salt crystal deposits (Table 20.1). In the deep-sea and subseafloor hydrothermal fluid systems, the rapid decompression of upwelling high-temperature hydrothermal fluids can induce phase separation and partition of the fluid into vapor- and brine-rich phases (Bischoff and Pitzer 1989). Phase-separation-influenced hydrothermal systems are well known, and both the highly brine-enriched and the vapor-dominating fluids have been identified (Von Damm 1995). It is thought that the brine-dominated fluids contribute to the high-saline subseafloor hydrothermal fluid flows. Furthermore, many deep-sea brine pools and deep subsurface salt deposits have been found (Table 20.1) (e.g., Swallow and Crease 1965; Krijgsman et al. 1999). These environments result from past evaporative events of seawater induced by sea level change and tectonic events. In contrast, environments with very low salt concentrations can be generated in the subseafloor environments by the inputs of terrestrial freshwater outflows near coasts, the condensation of vapor-phase hydrothermal fluids and magmatic volatiles, the dissociation water of gas (CO₂ and CH₄) hydrates that exclude the dissolved salts during hydrate formation, and the compressive extraction of the mineral formation water in the deep subsurface subduction zones. As the whole, however, the natural range of salinities in the Earth does not completely prevent the microbial growth and survival. Many freshwater microorganisms can grow in distilled water only supplemented with complex organic substrates, while extreme halophiles grow in NaCl-saturated media and can survive in salt crystals deposited in the deep subsurface environments over geologic time (Vreeland et al. 2000). The only cultivated extreme halophilic organism from the deep-sea and subseafloor high-salinity environments is an

extremely halophilic archaeon *Halorhabdus tiamatea*, which was isolated from deep-sea brine pool sediment in the Red Sea (Antunes et al. 2008), although many halophilic bacterial strains and prokaryotic 16S rRNA gene sequences have been identified in deep-sea hydrothermal vent chimneys and in deep-sea brine pools (Eder et al. 1999; Takai et al. 2001a; Antunes et al. 2011).

20.2.4 Nonionizing and Ionizing Radiations

Ultraviolet light such as UV-B (280–315 nm) and UV-C (100–280 nm) potentially affects the growth and survival of microbial communities in the surface environments due to the photochemical denaturation of functional biomolecules such as DNA and should have a great impact on propagation of life in the early Earth and the Universe. The maximum fluxes of UV-B and UV-C in the surface environments of the present Earth are estimated to be 2 W/m² and negligible, respectively (Cockell 2000). Of course, these UV-B and UV-C irradiances are not lethal and little inhibitory to most of organisms living in the surface environments. However, in the early (Hadean and early Archean) Earth environments, the surface UV flux is estimated to be much greater than in the present Earth, and not only UV-B but also UV-C reached the ground and ocean through the atmosphere due to the significantly different solar radiation and Earth's atmosphere composition, which would have affected the biological processes in the surface environments (Cockell 2000). The extent of biologically effective UV irradiance on the early Earth surface environments is predicted to be more than 10³ times greater (about 100 W/m²) than that on the present Earth (<0.1 W/m²) (Cockell 2000). In addition, the maximum fluxes of UV-B and UV-C in the thermosphere of Earth are measured by sensors on International Space Station (ISS) and are known to be 17.5 W/m² and 6.4 W/m², respectively (Table 20.1) (Schuster et al. 2012). These conditions are much more severe to life than the conditions on the present Earth surface and even in the early Earth surface environments.

Deinococcus spp. are known to be the most UV-irradiation-resistant microorganisms as well as the most ionizing-irradiation-resistant microorganisms (Table 20.2). Under the 2000 J/m² of UV-C irradiation, *Deinococcus radiodurans* shows 10⁻⁵ survival, and the dried states of cells are even more resistant (Battista 1997). With the maximum UV-irradiation-resistant ability, it is calculated that 10¹⁰ cells of *Deinococcus radiodurans* completely die in an hour in the early Earth surface environments and for several minutes in the ISS orbit when directly exposed to the light. Of course, the biologically detrimental UV irradiance is easily shielded by the existence of several tens meters depth of water and solid materials such as minerals and rocks and even the outer cellular and organic materials of cell aggregates (Cockell 2000; Horneck et al. 2001; Olsson-Francis and Cockell 2010). Thus, the UV radiation is not a decisive factor for restricting the microbial growth and survival in the subsurface environments and may not significantly affect the survival of endolithic microbial populations and cell aggregates in the outer space of Earth (see Chap. 27).

Ionizing radiation (gamma- and X-rays) may also affect the growth and survival of earthly life. The average dose rate of ionizing radiation on the surface environments is estimated to be below 1 mGy/year, while it becomes higher in high-altitude atmosphere and reaches to 430 mGy/year at the ground level in specific points of some high-radiation areas (Shahbazi-Gahrouei et al. 2013). However, certain subsurface environments have been found on Earth that are exposed to greater ionizing radiation doses than in the high-altitude atmosphere and the outer space. Microbial habitats closely located to the uranium mines on Earth are affected by the relatively high ionizing radiation doses (Gomez et al. 2006). In the past, the most extreme environments with ionizing radiation are expected to have occurred in the ancient natural fission reactors, as found in Oklo, Gabon (Table 20.1) (Jensen and Ewing 2001). Based on the alteration mineral chemistry, it is estimated that the environments several meters away from the reactor core were exposed to the temperature condition below 100 °C but >1 KGy/h of dose rate (Dartnell 2011). In this specific environment, even the most ionizing-irradiation-resistant *Deinococcus radiodurans*, which can revive maximally after 150 KGy dose of gamma-ray irradiation under the laboratory experimental conditions (Minton 1994; Battista 1997), would completely die in a week (Table 20.2). Other than in such spatially and temporally limited natural environments, ionizing radiation would have been not lethal and less inhibitory to life on the Earth.

Rather, it has been pointed out that nonionizing and ionizing radiations may have played significant roles in prebiotic chemical evolution of organic building block materials for birth of life (Dartnell 2011). On a related note, the deep-sea hydrothermal systems are among the naturally high-ionizing-radiation environments in the Earth (Charmasson et al. 2009). Indeed, on many of them, *Archaea* has demonstrated relatively high-ionizing-radiation resistance (Jolivet et al. 2004). Among them, *Thermococcus gammatolerans* is known to be the most ionizing radiation-resistant microorganism from the deep-sea and deep subsurface environments, which was isolated from the enrichment cultures after 30 KGy dose of gamma-ray irradiation (Jolivet et al. 2003).

20.2.5 Energetics

The microbial activity and survival in subsurface environments can be also limited by a lack of energy and essential elements (Valentine 2007). However, almost all the microbiological surveys of subsurface biospheres on the Earth have demonstrated that microorganisms are always present in the most energetically barren environments (Morono et al. 2012; D'Hondt et al. 2015). This is likely due to the fact that microorganisms, collectively, have tremendous metabolic potentials and can gain energy for growth, maintenance, and survival from numerous chemical redox reactions. It has been justified that the amount of energy and essential elements available in the habitat would control the biomass and function of microbial community and the emerging pattern of energy-transducing metabolic reactions could impact the

compositional and functional diversity of any microbial community (Takai and Nakamura 2011; Nakamura and Takai 2014, 2015; LaRowe and Amend 2015).

Not much is known of habitats where the energy and elemental fluxes are below that are needed to sustain a detectable microbial community; a theoretical formulation has provided an energy balance concept of habitability that may establish the energetic basis of limits of life and biosphere in the Earth and even in the Universe (Fig. 20.1) (Hoehler 2007). The concept suggests that the habitability occurs when the potential for transduction of environmental energy into biological process outweighs the biological demand for energy, and the biological demand for energy is a two-dimensional (power and voltage) function that encompasses biochemical specifics, physical and chemical conditions, and material or solvent limitations (Fig. 20.1) (Hoehler 2007). In addition, it is also revealed that habitability is determined by both voltage and power uptake limits (Fig. 20.1) (Hoehler 2007). This energy balance concept is applicable not only to the habitability on Earth but also elsewhere, indicating that the habitable or uninhabitable states of environment could be to large extent diagnosed based on the physical and chemical conditions and the material or solvent availability even though the biochemical specificity is highly uncertain. For the earthly life, however, the biological demand minimums for energy and the voltage and power uptake limits have been empirically and theoretically characterized via laboratory experiments and field observations. The presently estimated values that could determine the energetic habitability of a unicellular earthly life are introduced in the following paragraphs (Fig. 20.1).

It was long believed that about 60 KJ free energy change is required for 1 mol ATP synthesis by the respiratory ATP synthase usually requiring 3 mol H^+ across the membrane. Thus about -20 KJ/mol e^- is a certain energy quantum for ATP synthesis, more so, an energy quantum for life (Schink 1997). However, Jackson and

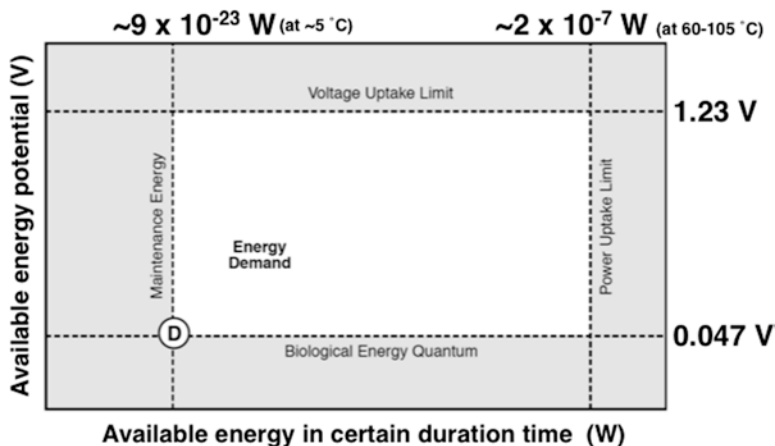


Fig. 20.1 Energy balance concept of habitability modified from the original one by Hoehler (2007). The biological energy quantum, the maintenance energy, and the voltage and power uptake limits are calculated in this chapter

McInerney (2002) found that microbial syntrophic growth, particularly the growth of its fermentative bacterial component, requires a free energy change as low as -4.5 KJ/mol e^- , which is equivalent to 0.047 V of redox potential change. This is the minimum free energy change in the whole energy transduction reaction during the microbial growth or maintenance ever observed and represents the biological energy quantum for earthly life that Hoehler (2007) pointed to as one dimension of the biological demand for energy (Fig. 20.1). In contrast, the voltage uptake limit for earthly life is not empirically and experimentally determined regardless of the fact that the upper limit is theoretically fixed to the standard potential (1.23 V) of electrolysis of water (Fig. 20.1). This implies that 1.23 V and greater of redox potential change would induce the electrolysis of cellular water and lead to fatal impacts on the growth, maintenance, and survival of earthly life and even extraterrestrial life using water as the living solvent.

The maintenance energy, the other dimension of the biological demand for energy, is defined as the minimum energy requirement to sustain life in certain duration of time. The minimum energy requirement to sustain life has been estimated based on the laboratory experiments (Heijnen and van Dijken 1992) and thermodynamic calculation (McCollom and Amend 2005) of biomass reproduction for various microorganisms and metabolisms. Under anaerobic conditions, the minimum energy requirement to sustain life is the least for H_2 -utilizing (hydrogenotrophic) chemoautotrophs and is found to be $30\text{--}40 \text{ KJ/g}$ dry cell by laboratory experiments (Heijnen and van Dijken 1992) and is estimated to be 1.4 KJ/g dry cell by thermodynamic calculation (McCollom and Amend 2005). Given that the smallest cell of a hydrogenotrophic chemoautotroph has 10 fg dry cellular carbon and 20 fg dry weight (Fukuda et al. 1998), the value of 1.4 KJ/g dry cell means that $2.8 \times 10^{-11} \text{ J}$ of energy is required to sustain one smallest body of earthly life. In addition, the reproduction rate of the smallest cell using the least maintenance energy should be at least faster than the denaturation rates of the most labile cellular components, for example, which is represented by the racemization rate of aspartate in proteins, primarily dependent on temperature (Price and Sowers 2004). At $5 \text{ }^\circ\text{C}$, as the racemization of aspartate likely reaches to the equilibrium for about 10^4 years, the smallest cell of life should reproduce itself at least in 10^4 years. According to these values, the theoretical minimum of maintenance energy is estimated to be $9 \times 10^{-23} \text{ W}$ under the temperature condition of $5 \text{ }^\circ\text{C}$ (Fig. 20.1). As described above, it should be noted that the minimum of maintenance energy is primarily dependent on the temperature condition of the environment.

Finally, the power uptake limit for earthly life is the most difficult to determine, and there has been no estimation reported so far. It may be possible to estimate an approximated value from the maximum intracellular ATP amount (convertible into the maximum intracellular energy stock) of a microbial cell under abundantly energy- and nutrient-enriched conditions. This means the intracellular maximum energy stock potential. The maximum intracellular ATP amount is theoretically calculated to be 0.045 mol/g dry cell for *E. coli* under the ideal optimal condition for growth, implying the condition under which the most efficient balance of intracellular ATP synthesis and consumption is attained and the highest concentration of

ATP is pooled in the cell for the shortest time of 20 min (Russell and Cook 1995). According to a value of 20 fg dry cell described above, the maximum intracellular ATP equilibrium concentration in the cell is 0.9×10^{-15} mol, which corresponds to 2.9×10^{-14} W of power stored in the smallest body of earthly life when the whole ATP is consumed as energy transduction source. Another estimation for power uptake limit may be obtained from the greatest energy consumption of the fastest growth of unicellular earthly life under the abundantly energy- and nutrient-enriched conditions. This means the maximum energy consumption potential of microbial cell. As far as known, the shortest doubling time of the fastest-growing bacterium (an anaerobic fermentative thermophilic bacterium) is reported to be 10 min at 60 °C (Elsgaard and Prieur 2011). Since the energy consumption of this bacterium during the fastest growth was not estimated, the greatest energy consumption of fast-growing microorganism is calculated, for instance, from the case of optimal growth for *Methanopyrus kandleri* strain 116 (Takai et al. 2008a). During the optimal growth under a highly H₂-abundant and piezophilic condition (at 105 °C and 20 MPa), the methanogenic archaeon maximally consumes 3.7×10^{-9} M of H₂ (9.3×10^{-10} M CH₄ production) per cell that is equivalent to 1.2×10^{-4} J of energy consumption per cell. Assuming that the similar energy consumption would occur in the fastest-growing bacterium, the maximum energy consumption potential of a microbial cell is calculated to be 2.0×10^{-7} W. This value represents an estimated power uptake limit for the known earthly life (Fig. 20.1).

20.3 Challenge for Limits of Biosphere in the Subsurface Environments

Several scientific drilling expeditions have challenged to explore the realistic upper temperature limits for life and biosphere in the seafloor environments beneath the hydrothermal vent systems (Cragg and Parkes 1994; Reysenbach et al. 1998; Cragg et al. 2000; Kimura et al. 2003; Takai et al. 2011) and non-hydrothermal areas with relatively large heat flow (Blackman et al. 2006; Heuer et al. 2017). For example, the Leg 158 of the Ocean Drilling Program (ODP) was conducted in the proximity of high-temperature hydrothermal discharges at the TAG field in the Mid-Atlantic Ridge (Humphris et al. 1996). The ODP Legs 139 and 169 investigated the hydrothermal vent and flank regions of the sediment-covered hydrothermal system at the Middle Valley field on the Juan de Fuca Ridge (Davis et al. 1992; Fouquet et al. 1998). The ODP Leg 193 particularly explored the sulfide deposits close to the active hydrothermal vents at the PACMANUS field in the Manus Basin (Binns et al. 2002), and the Integrated Ocean Drilling Program (IODP) Expedition 331 investigated the hydrothermal activity center near the Iheya North field in the Okinawa Trough (Takai et al. 2011). In addition, the IODP Expedition 304/305 obtained plutonic rock samples from the crest of Southern Ridge in the Atlantis Massif (Blackman et al. 2006), which hosts the serpentinization-driven Lost City hydrothermal vent

field (about 4 km south from Site U1309) (Kelley et al. 2001, 2005), and the IODP Expedition 370 obtained the sedimentary core samples in the Nankai Trough off the Cape Muroto (Heuer et al. 2017).

A number of microbiological techniques were used to search for microorganisms in core samples taken at various depths of 0–52.1 mbsf (meters below sea floor) during ODP Leg 158 (Humphris et al. 1996). Microscopy, attempts at cultivation, and DNA extractions all failed to find evidence of a biosphere or viable microbial communities in the subsurface near the TAG hydrothermal field (Reysenbach et al. 1998). However, it was not clear whether the environment beneath the TAG field is uninhabitable for life or the low microbial populations were undetectable due to the inherent technical and methodological limitations.

In the adjacent and flank regions of the Middle Valley hydrothermal field on the Juan de Fuca Ridge, microbial cell counts were carried out on sediment core samples that were influenced by seafloor hydrothermal fluid flow (Cragg and Parkes 1994; Cragg et al. 2000). The vertical distribution of microbial communities in the area with the steep thermal gradients (Site 858) was detected at down to a depth of 67 mbsf (Cragg and Parkes 1994). Estimates of the temperature constraints suggested that the potentially viable microbial communities were restricted to relatively shallow seafloor environments which were lower than 76 °C (Cragg and Parkes 1994). However, in several deeper seafloor horizons, probably where the hydrothermal fluid circulates, detectable microbial populations were observed at a temperature range of 155–185 °C (Cragg and Parkes 1994; Cragg et al. 2000). Similar depth and temperature profiles of microbial cell abundance were obtained from sediments in the hydrothermal flank regions (Site 1036) (Cragg et al. 2000). However, the viability of these cells, their functions, and their origins have not been determined, since they are simply counted based on acridine orange stain.

The ocean drilling expedition in the PACMANUS hydrothermal field (Leg 193) was the first to be conducted in an active arc-back-arc hydrothermal system (Binns et al. 2002). Coring operations extended down to 387 mbsf, and samples were subjected to microbiological characterization (Kimura et al. 2003). Microbial cell populations were detected in the core samples at depths down to about 69–80 mbsf (Sites 1188 and 1189), while the ATP concentrations indicative of viable microbial populations were quantitatively significant down to 39–49 mbsf (Sites 1188 and 1189) (Kimura et al. 2003). Although the bottom temperatures of the drilled holes were determined by using wire-line logging tools several days after the drilling operation, the temperature ranges of detectable microbial cell populations and potentially viable microbial populations were uncertain, since they were only based on the bottom temperature measurements (Kimura et al. 2003). Perhaps, potentially viable microbial communities are more extensive at relatively low temperatures (shallower zones) little affected by seafloor high-temperature hydrothermal fluid flow. In addition, successful enrichments of thermophiles were obtained from the core samples, some of which were even deeper than the depth limit of detectable microbial cell populations (Kimura et al. 2003). It is still uncertain whether these thermophiles derived from the indigenous viable or survived microbial communities or from drilling fluids that were contaminated at shallower depths.

These ODP expedition-based microbiological investigations for the upper temperature limits of biosphere have provided some operational and analytic guidelines for the comprehensive investigations. With the guidelines, a biogeochemistry- and microbiology-dedicated IODP expedition (IODP Exp. 331) was conducted in the Iheya North hydrothermal field in the middle Okinawa Trough (Takai et al. 2011; Yanagawa et al. 2017). The onboard microbial cell counts revealed that the microbial cell populations were likely distributed in the hydrothermally active seafloor sediments down to a depth of 15–20 mbsf (for one drilling site) (Takai et al. 2011; Yanagawa et al. 2017). In the sediments, the in situ temperatures were measured to be about 45 °C at the time of drilling by several temperature monitors but were estimated to be >106 °C before the drilling based on the geochemical thermometers using hydrothermal alteration minerals (Yanagawa et al. 2017). Multiple lines of evidence for the possible distribution of viable microbial communities were obtained from various biogeochemical and microbiological characterizations, such as stable isotope analyses of various energy, carbon, nitrogen, and sulfur sources, prokaryotic 16S rRNA and functional gene detection and quantification, metabolic activity measurements, and cultivation tests (Aoyama et al. 2014; Yanagawa et al. 2017). All the data suggest that the possible boundary between the habitable and the uninhabitable regions are present at the depth of 15–20 mbsf under the formation temperature of >106 °C (Yanagawa et al. 2017). This study represents the first example of the field observations that determined the realistic limit of biosphere and the boundary temperature condition between the habitable and uninhabitable terrains in the subsurface environments.

The IODP expedition 304/305 explored the plutonic rock environments with relatively steep thermal gradients in the Atlantis Massif near the MAR (Blackman et al. 2006). One of the drilling holes penetrated down to 1415.5 mbsf through very thin sediments and massive gabbroic rocks and obtained core samples down to 1395 mbsf (Blackman et al. 2006). After the completion of drilling, a temperature measurement was performed throughout the drilled depth, and the bottom of the hole at 1415 mbsf was found to be 119 °C (Blackman et al. 2006). For the purpose of microbiological investigation, 26 sediment and rock samples were subsampled from the cores at depths of 0.45–1391 mbsf in order to carry out microbial cell counts and PCR-based 16S rRNA gene analysis (Mason et al. 2010). Except for the uppermost carbonate sediments (0.5 mbsf), no evidence of microbial cellular population was detected. However, 16S rRNA genes were amplified from DNA extracted from the gabbroic rocks (Mason et al. 2010). The maximal temperature for the bacterial 16S rRNA gene detection was 79 °C at a depth of 1313 mbsf (Mason et al. 2010). However, the bacterial phylotypes obtained from ~80 °C of rock habitats were closely related with mesophilic *Ralstonia* spp., indicating that microbial contamination was likely (Mason et al. 2010).

Very recently, IODP Expedition 370 has explored the temperature limit of seafloor sedimentary microbial community development with a relative steep thermal gradient in the Nankai Trough off the Cape Muroto (Heuer et al. 2017). Although the detailed biogeochemical and microbiological investigations are still ongoing, the preliminary onboard and onshore results have suggested that the

microbial cell populations are detectable at up to the in situ temperature condition of 60 °C (Heuer et al. 2017). Interestingly, the temperature condition is similar to the apparent temperature limits of possible seafloor microbial cell populations in Site 858 of the Middle Valley field (Cragg and Parkes 1994) and putative indigenous microbial cell populations and prokaryotic 16S rRNA gene sequences in the deep continental margin sediments (60–100 °C) during ODP and IODP expeditions (Roussel et al. 2008; Ciobanu et al. 2014). However, all these temperature conditions still stand on the large uncertainty of partial observation and analysis for the subsurface microbial community development. The multiple lines of evidences for an upper temperature limit of biosphere and the reliable boundary temperature condition (>106 °C) in the subsurface environments are successfully retrieved only from the Iheya North hydrothermal field in the middle Okinawa Trough (Takai et al. 2011; Yanagawa et al. 2017).

In the geomicrobiological exploration of deep buried marine sediments off the Shimokita Peninsula, the multiple boundaries between the presence and absence of indigenous microbial cell populations were clearly identified under the temperature condition of 40–60 °C in below 1.5 kmbf (Inagaki et al. 2015). The realistic limits of biosphere were based on the reliable observation and analysis, while the temperature would be not the only decisive factor to determine the boundaries because the habitable and uninhabitable zones coexisted in the same temperature range (Inagaki et al. 2015). Based on the physical and chemical conditions of these habitable and uninhabitable zones of sediments, the potential balance of temperature-dependent biological income (catabolic) and outgo (maintenance) energy was thermodynamically estimated. As the result, it seemed very likely that the temperature-dependent energetic balance becomes the boundary conditions of limits of biosphere in the deep buried marine sediments (Inagaki et al. 2015). This is the first example showing that the energy balance concept of habitability proposed by Hoehler (2007) reflects the realistic limits of biosphere in natural environments.

20.4 Concluding Remarks and Perspectives

In this chapter, many potential physical and chemical constraints to determine realistic limits of earthly life and biosphere were described. Significant gaps between the results obtained from the laboratory experiments and the field observations have been observed. The natural environments have the great spatial and temporal fluctuation of physical and chemical constraints and host the heterogeneous microbial populations and metabolisms based on the symbiotic and synergetic interactions. Since most of these physical and chemical constraints and biological interactions directly or indirectly affect the energetics of metabolisms for growth, maintenance, and survival of life, thermodynamic estimations have become a very powerful approach to determine in which environments particular catabolic strategies are favored and how much cost is required for growth, maintenance, and survival of life in different extreme environments. Based on these kinds of thermodynamic

calculations, the future explorations for realistic limits of life and biosphere in the present Earth can be rationalized. In a similar manner, the occurrence and propagation of life and biosphere in the early Earth's environments can be probabilistically addressed, and the search and detection of extraterrestrial life will be better planned and conducted in the future.

References

- Antunes A, Taborda M, Huber R, Moissl C, Nobre MF, da Costa MS (2008) *Halorhabdus tiamatea* sp. nov., a non-pigmented, extremely halophilic archaeon from a deep-sea, hypersaline anoxic basin of the Red Sea, and emended description of the genus *Halorhabdus*. *Int J Syst Evol Microbiol* 58:215–220
- Antunes A, Ngugi DK, Stingl U (2011) Microbiology of the Red Sea (and other) deep-sea anoxic brine lakes. *Environ Microbiol Rep* 3:416–433
- Aoyama S, Nishizawa M, Takai K, Ueno Y (2014) Microbial sulfate reduction within the Iheya North subseafloor hydrothermal system constrained by quadruple sulfur isotopes. *Earth Planet Sci Lett* 398:113–126
- Aoyama S, Nishizawa M, Miyazaki J, Shibuya T, Ueno Y, Takai K (2018) Recycled Archean sulfur in the mantle wedge of the Mariana Forearc and microbial sulfate reduction within an extremely alkaline serpentine seamount. *Earth Planet Sci Lett* 491:109–120
- Battista JR (1997) Against all odds: the survival strategies of *Deinococcus radiodurans*. *Ann Rev Microbiol* 51:203–224
- Belliaev GM, Brueggeman PL (1989) Deep sea ocean trenches and their Fauna. Scripps Institution of Oceanography Technical Report, Scripps Institution of Oceanography, UC San Diego
- Binns RA, Barriga FJAS, Miller DJ, Shipboard Scientific Party (2002) Proc ODP Init Rep 193. doi:<https://doi.org/10.2973/odp.proc.ir.193.2002>
- Bischoff JL, Pitzer KS (1989) Liquid-vapor relations for the system NaCl-H₂O; summary of the P-T-x surface from 300 degrees to 500 degrees C. *Am J Sci* 289:217–248
- Bischoff JL, Rosenbauer RJ (1988) Liquid-vapor relations in the critical region of the system NaCl-H₂O from 380 to 415 °C: a refined determination of the critical point and two-phase boundary of seawater. *Geochim Cosmochim Acta* 52:2121–2126
- Blackman DK, Ildefonse B, John BE, Ohara Y, Miller DJ, MacLeod CJ, The Expedition 304/305 Scientists (2006) Proc IODP Exp 304/305. doi:<https://doi.org/10.2204/iodp.proc.304305.101.2006>
- Butterfield DA, Nakamura K, Takano B, Lilley MD, Lupton JE, Resing JA, Roe KK (2011) High SO₂ flux, sulfur accumulation, and gas fractionation at an erupting submarine volcano. *Geology* 39:803–806
- Charmasson S, Sarradin PM, Le Faouder A, Agarande M, Loyer J, Desbruyeres D (2009) High levels of natural radioactivity in biota from deep-sea hydrothermal vents: a preliminary communication. *J Environ Radioact* 100:522–526
- Ciobanu MC, Burgaud G, Dufresne A, Breuker A, Redou V, Maamar SB, Gaboyer F, Vandenabeele-Trambouze O, Lipp JS, Schippers A, Vandenkoornhuyse P, Barbier G, Jebbar M, Godfroy A, Alain K (2014) Microorganisms persist at record depths in the subseafloor of the Canterbury Basin. *ISME J* 8:1370–1380
- Cockell CS (2000) Ultraviolet radiation and the photobiology of Earth's early oceans. *Orig Life Evol Biosph* 30:467–500
- Cragg BA, Parkes RJ (1994) Bacterial profiles in hydrothermally active deep sediment layers from Middle Valley (NE Pacific), Sites 857 and 858. *Proc ODP Sci Results* 139:509–516
- Cragg BA, Summit M, Parkes RJ (2000) Bacterial profiles in a sulfide mound (Site 1035) and an area of active fluid venting (Site 1036) in hot hydrothermal sediments from Middle

- Valley (Northeast Pacific). Proc ODP Sci Result 169. doi:<https://doi.org/10.2973/odp.proc.sr.169.105.2000>
- D'Elia T, Veerapaneni R, Rogers SO (2008) Isolation of microbes from Lake Vostok accretion ice. Appl Environ Microbiol 74:4962–4965
- D'Hondt S, Inagaki F, Zarikian CA, Abrams LJ, Dubois N, Engelhardt T, Evans H, Ferdelman T, Gribsholt B, Harris R, Hoppie BW, Hyun JH, Kallmeyer J, Kim J, Lynch JE, McKinley CC, Mitsunobu S, Morono Y, Murray RW, Pockalny R, Sauvage J, Shimono T, Shiraishi F, Smith DC, Smith-Duque CE, Spivack AJ, Steinsbu BO, Suzuki Y, Szpak M, Toffin L, Uramoto G, Yamaguchi Y, Zhang GL, Zhang XH, Ziebis W (2015) Presence of oxygen and aerobic communities from sea floor to basement in deep-sea sediments. Nat Geosci 8:299–304
- Dalmasso C, Oger P, Selva G, Courtine D, L'Haridon S, Garlaschelli A, Roussel E, Miyazaki J, Reveillaud J, Jebbar M, Takai K, Maignien L, Alain K (2016) *Thermococcus piezophilus* sp. nov., a novel hyperthermophilic and piezophilic archaeon with a broad pressure range for growth, isolated from a deepest hydrothermal vent at the Mid-Cayman Rise. Syst Appl Microbiol 39:440–444
- Darrah TH, Tedesco D, Tassi F, Vaselli O, Cuoco E, Poreda RJ (2013) Gas chemistry of the Dallol region of the Danakil Depression in the Afar region of the northern-most East African Rift. Chem Geol 339:16–29
- Dartnell LR (2011) Ionizing radiation and life. Astrobiology 11:551–582
- Davis EE, Mottl MJ, Fisher AT, Shipboard Scientific Party (1992) Proc ODP Init Rep 139. doi:<https://doi.org/10.2973/odp.proc.ir.139.1992>
- Doemel WN, Brock TD (1971) The physiological ecology of *Cyanidium caldarium*. J Gen Microbiol 67:17–32
- Eder W, Ludwig W, Huber R (1999) Novel 16S rRNA gene sequences retrieved from highly saline brine sediments of Kebrit Deep, Red Sea. Arch Microbiol 172:213–218
- Edwards KJ, Bond PL, Gihring TM, Banfield JF (2000) An archaeal iron-oxidizing extreme acidophile important in acid mine drainage. Science 287:1796–1799
- Elsgaard L, Prieur D (2011) Hydrothermal vents in Lake Tanganyika harbor spore-forming thermophiles with extremely rapid growth. J Great Lakes Res 37:203–206
- Fouquet Y, Zierenberg RA, Miller DJ, Shipboard Scientific Party (1998) Proc ODP Init Rep 169. doi:<https://doi.org/10.2973/odp.proc.ir.169.1998>
- Fukuda R, Ogawa H, Nagata T, Koike I (1998) Direct determination of carbon and nitrogen contents of natural bacterial assemblages in marine environments. Appl Environ Microbiol 64:3352–3358
- Glud RN, Wenzhöfer F, Middelboe M, Oguri K, Turnewitsch R, Canfield DE, Kitazato H (2013) High rates of microbial carbon turnover in sediments in the deepest oceanic trench on Earth. Nat Geosci 6:284–288
- Gomez P, Garralon A, Buil B, Turrero MJ, Sanchez L, de la Cruz B (2006) Modeling of geochemical processes related to uranium mobilization in the groundwater of a uranium mine. Sci Total Environ 366:295–309
- Heijnen JJ, van Dijken JP (1992) In search of a thermodynamic description of biomass yield for the chemotrophic growth of microorganisms. Biotechnol Bioeng 39:833–858
- Heuer VB, Inagaki F, Morono Y, Kubo Y, Maeda L, the IODP Exp 370 Onboard scientists (2017) Expedition 370 preliminary report: temperature limit of the deep biosphere off Muroto. International Ocean Discovery Program. doi:<https://doi.org/10.14379/iodp.pr.370.2017>
- Hoehler TM (2007) An energy balance concept for habitability. Astrobiology 7:824–838
- Horikoshi K (1999) Alkaliphiles: some applications of their products for biotechnology. Microbiol Mol Biol Rev 63:735–750
- Horneck G, Rettberg P, Reitz G, Wehner J, Eschweiler U, Strauch K, Verena CP, Baumstark-Khan C (2001) Protection of bacterial spores in space, a contribution to the discussion on panspermia. Orig Life Evol Biosph 31:527–547
- Humphris SE, Herzig PM, Miller DJ, Shipboard Scientific Party (1996) Proc ODP Init Rep 158. doi:<https://doi.org/10.2973/odp.proc.ir.158.1996>

- Imshenetsky AA, Lysenko SV, Kazakov GA, Ramkova NV (1976) On micro-organisms of the stratosphere. *Life Sci Space Res* 14:359–362
- Inagaki F, Hinrichs KU, Kubo Y, Bowles MW, Heuer VB, Hong WL, Hoshino T, Ijiri A, Imachi H, Ito M, Kaneko M, Lever M, Lin YS, Methé BA, Morita S, Morono Y, Tanikawa W, Bihan M, Bowden S, Elvert M, Glombitza C, Gross D, Harrington GJ, Hori T, Li K, Limmer D, Liu CH, Murayama M, Ohkouchi N, Ono S, Park YS, Phillips SC, Prieto-Mollar X, Purkey M, Riedinger N, Sanada Y, Sauvage J, Snyder G, Susilawati R, Takano Y, Tasumi E, Terada T, Tomaru H, Trembath-Reichert E, Wang DT, Yamada Y (2015) Exploring deep microbial life in coal-bearing sediment down to ~2.5 km below the ocean floor. *Science* 349:420–424
- Jackson BE, McInerney MJ (2002) Anaerobic microbial metabolism can proceed close to thermodynamic limits. *Nature* 415:454–456
- Jamieson AJ, Malkocs T, Piertney SB, Fujii T, Zhang Z (2017) Bioaccumulation of persistent organic pollutants in the deepest ocean fauna. *Nature Ecol Evol* 1:0051
- Jensen KA, Ewing RC (2001) The Okélobondo natural fission reactor, southeast Gabon: geology, mineralogy, and retardation of nuclear-reaction products. *GSA Bull* 113:32–62
- Jolivet E, Corre E, L'Haridon S, Forterre P, Prieur D (2003) *Thermococcus gammatolerans* sp. nov., a hyperthermophilic archaeon from a deep-sea hydrothermal vent that resists ionizing radiation. *Int J Syst Evol Microbiol* 53:847–851
- Jolivet E, Corre E, L'Haridon S, Forterre P, Prieur D (2004) *Thermococcus marinus* sp. nov. and *Thermococcus radiotolerans* sp. nov., two hyperthermophilic archaea from deep-sea hydrothermal vents that resist ionizing radiation. *Extremophiles* 8:219–227
- Kato C, Li L, Tamaoka J, Horikoshi K (1997) Molecular analyses of the sediment of the 11,000-m deep Mariana Trench. *Extremophiles* 1:117–123
- Kawagucci S, Miyazaki J, Morono Y, Seewald JS, Wheat G, Takai K (2018) Cool and alkaline serpentinite formation fluid regime with scarce microbial habitability and possible abiotic synthesis beneath the South Chamorro Seamount. *Proc Earth Planet Sci* 5:1–20
- Kelley DS, Karson JA, Blackman DK, Fruh-Green GL, Butterfield DA, Lilley MD, Olson EJ, Schrenk MO, Roe KK, Lebon GG, Rivizzigno P (2001) An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30 degrees N. *Nature* 412:145–149
- Kelley DS, Karson JA, Fruh-Green GL, Yoerger DR, Shank TM, Butterfield DA, Hayes JM, Schrenk MO, Olson EJ, Proskurowski G, Jakuba M, Bradley A, Larson B, Ludwig K, Glickson D, Buckman K, Bradley AS, Brazelton WJ, Roe K, Elend MJ, Delacour A, Bernasconi SM, Lilley MD, Baross JA, Summons RE, Sylva SP (2005) A serpentinite-hosted ecosystem: the lost city hydrothermal field. *Science* 307:1428–1434
- Kimura H, Asada R, Masta A, Naganuma T (2003) Distribution of microorganisms in the subsurface of the Manus basin hydrothermal vent field in Papua New Guinea. *Appl Environ Microbiol* 69:644–648
- Klotz B, Pyle DL, Mackey BM (2007) New mathematical modeling approach for predicting microbial inactivation by high hydrostatic pressure. *Appl Environ Microbiol* 73:2468–2478
- Kobayashi H, Hatada Y, Tsubouchi T, Nagahama T, Takami H (2012) The hadal amphipod *Hirondellea gigas* possessing a unique cellulase for digesting wooden debris buried in the deepest seafloor. *PLoS One* 7:e42727
- Koschinsky A, Garbe-Schönberg D, Sander S, Schmidt K, Gennerich HH, Strauss H (2008) Hydrothermal venting at pressure-temperature conditions above the critical point of seawater, 5°S on the Mid-Atlantic Ridge. *Geology* 36:615–618
- Krijgsman W, Hilgen FJ, Raffi I, Sierro FJ, Wilson DS (1999) Chronology, causes and progression of the Messinian salinity crisis. *Nature* 400:652–655
- Kusube M, Kyaw TS, Tanikawa K, Chastain RA, Hardy KM, Cameron J, Bartlett DH (2017) *Colwellia marinimaniae* sp. nov., a hyperpiezophilic species isolated from an amphipod within the Challenger Deep, Mariana trench. *Int J Syst Evol Microbiol* 67:824–831
- LaRowe DE, Amend JP (2015) Catabolic rates, population sizes and doubling/replacement times of microorganisms in natural settings. *Am J Sci* 315:167–203

- Margosch D, Ehrmann MA, Buckow R, Heinz V, Vogel RF, Gänzle MG (2006) High-pressure-mediated survival of *Clostridium botulinum* and *Bacillus amyloliquefaciens* endospores at high temperature. *Appl Environ Microbiol* 72:3476–3481
- Mason OU, Nakagawa T, Rosner M, Van Nostrand JD, Zhou J, Maruyama A, Fisk MR, Giovannoni SJ (2010) First investigation of the microbiology of the deepest layer of ocean crust. *PLoS One* 5:e15399
- McCollom TM, Amend JP (2005) A thermodynamic assessment of energy requirements for biomass synthesis by chemolithoautotrophic micro-organisms in oxic and anoxic environments. *Geobiology* 3:135–144
- McCollom TM, Bach W (2009) Thermodynamic constraints on hydrogen generation during serpentinization of ultramafic rocks. *Geochim Cosmochim Acta* 73:856–875
- Méndez-García C, Peláez AI, Mesa V, Sánchez J, Golyshina OV, Ferrer M (2015) Microbial diversity and metabolic networks in acid mine drainage habitats. *Front Microbiol* 6:475
- Minton KW (1994) DNA repair in the extremely radioresistant bacterium *Deinococcus radiodurans*. *Mol Microbiol* 13:9–15
- Morono Y, Terada T, Nishizawa M, Ito M, Hillion F, Takahata N, Sano Y, Inagaki F (2012) Carbon and nitrogen assimilation in deep seafloor microbial cells. *Proc Natl Acad Sci U S A* 108:18295–18300
- Mottl MJ (2009) Highest pH. *Geochem News* 141:09
- Mottl MJ, Komor SC, Fryer P, Moyer CL (2003) Deep-slab fluids fuel extremophilic Archaea on a Mariana forearc serpentinite mud volcano: ocean Drilling Program Leg 195. *Geochem Geophys Geosyst* 4:11
- Nakagawa T, Takai K, Suzuki Y, Hirayama H, Konno U, Tsunogai U, Horikoshi K (2006) Geomicrobiological exploration and characterization of a novel deep-sea hydrothermal system at the TOTO caldera in the Mariana Volcanic Arc. *Environ Microbiol* 8:37–49
- Nakamura K, Takai K (2014) Theoretical constraints of physical and chemical properties of hydrothermal fluids on variations in chemolithotrophic microbial communities in seafloor hydrothermal systems. *Prog Earth Planet Sci* 1:5
- Nakamura K, Takai K (2015) Geochemical constraints on potential biomass sustained by seafloor water–rock interactions. In: Ishibashi J, Okino K, Sunamura M (eds) *Seafloor biosphere linked to hydrothermal systems*. Springer Japan, Tokyo, pp 11–30
- Nordstrom DK, Alpers CN, Ptacek CJ, Blowes DW (2000) Negative pH and extremely acidic mine waters from Iron Mountain, California. *Environ Sci Tech* 34:254–258
- Nunoura T, Takaki Y, Hirai M, Shimamura S, Makabe A, Koide O, Kikuchi T, Miyazaki J, Koba K, Yoshida N, Sunamura M, Takai K (2015) Hadal biosphere: insight into the microbial ecosystem in the deepest ocean on Earth. *Proc Natl Acad Sci U S A* 112:E1230–E1236
- Olsson-Francis K, Cockell CS (2010) Experimental methods for studying microbial survival in extraterrestrial environments. *J Microbiol Methods* 80:1–13
- Pedersen K, Nilsson E, Arlinger J, Hallbeck L, O'Neill A (2004) Distribution, diversity and activity of microorganisms in the hyper-alkaline spring waters of Maqarin in Jordan. *Extremophiles* 8:151–164
- Price PB, Sowers T (2004) Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. *Proc Natl Acad Sci U S A* 101:4631–4636
- Rainy FA, Oren A (2006) Extremophile microorganisms and the method to handle them, *Method Microbiol* vol 35 *Extremophiles*. Elsevier, London
- Resing JA, Lebon G, Baker ET, Lupton JE, Embley RW, Massoth GJ, Chadwick WW, de Ronde CEJ (2007) Venting of acid-sulfate fluids in a high-sulfidation setting at NW rota-1 submarine volcano on the Mariana Arc. *Econ Geol* 102:1047–1061
- Reysenbach AL, Holm NG, Hershberger K, Prieur D, Jeanthou C (1998) In search of a subsurface biosphere at a slow-spreading ridge. *Proc ODP Sci Results* 158:355–365
- Reysenbach AL, Liu Y, Banta AB, Beveridge TJ, Kirshtein JD, Schouten S, Tivey MK, Vom Domm KL, Voytek MA (2006) A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. *Nature* 442:444–447

- Rothschild LJ, Mancinelli RL (2001) Life in extreme environments. *Nature* 409:1092–1101
- Roussel EG, Cambon Bonavita MA, Querellou J, Cragg BA, Webster G, Prieur D, Parkes RJ (2008) Extending the sub-sea-floor biosphere. *Science* 320:1046
- Rowe GL, Brantley SL, Fernandez M, Fernandez JF, Borgia A, Barquero J (1992) Fluid-volcano interaction in an active stratovolcano: the crater lake system of Poás volcano, Costa Rica. *J Volcanol Geotherm Res* 49:23–51
- Russell JB, Cook GM (1995) Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol Rev* 59:48–62
- Salisbury MH, Shinohara M, Richter C, Shipboard Scientific Party (2002) Proc ODP Init Rep 195. doi:<https://doi.org/10.2973/odp.proc.ir.195.2002>
- Schink B (1997) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol Mol Biol Rev* 61:262–280
- Schleper C, Pühler G, Kühlmorgen B, Zillig W (1995) Life at extremely low pH. *Nature* 375:741–742
- Schrenk MO, Brazelton WJ, Lang SQ (2013) Serpentinization, carbon, and deep life. *Rev Mineral Geochem* 75:575–606
- Schuster M, Dachev T, Richter P, Häder DP (2012) R3DE: radiation risk radiometer-dosimeter on the International Space Station—optical radiation data recorded during 18 months of EXPOSE-E exposure to open space. *Astrobiology* 12:393–402
- Seyfried WE Jr, Ding K, Berndt ME (1991) Phase equilibria constraints on the chemistry of hot spring fluids at mid-ocean ridges. *Geochim Cosmochim Acta* 55:3559–3580
- Shahbazi-Gahrouei D, Gholami M, Setayandeh S (2013) A review on natural background radiation. *Adv Biomed Res* 2:65
- Shock EL (1992) Chemical environments of submarine hydrothermal systems. *Orig Life Evol Biosph* 22:67–107
- Spies FN, RISE Group (1980) East Pacific Rise; hot springs and geophysical experiments. *Science* 207:1421–1433
- Swallow JC, Crease J (1965) Hot salty water at the bottom of the Red Sea. *Nature* 205:165–166
- Takai K (2011) Limits of life and the biosphere: lessons from the detection of microorganisms in the deep-sea and deep subsurface of the Earth. In: Gargaud M, Lopez-Garcia P, Martin H (eds) *Origins and evolution of life – an astrobiological perspective*. Cambridge University Press, Cambridge, UK, pp 469–486
- Takai K, Nakamura K (2011) Archaeal diversity and community development in deep-sea hydrothermal vents. *Curr Opin Microbiol* 14:282–291
- Takai K, Inoue A, Horikoshi K (1999) *Thermaerobacter marianensis* gen. nov., sp. nov., an aerobic extremely thermophilic marine bacterium from the 11,000 m deep Mariana Trench. *Int J Syst Bacteriol* 49:619–628
- Takai K, Komatsu T, Inagaki F, Horikoshi K (2001a) Distribution and colonization of archaea in a black smoker chimney structure. *Appl Environ Microbiol* 67:3618–3629
- Takai K, Moser DP, Onstott TC, Spoelstra N, Piffner SM, Dohnalkova A, Fredrickson JK (2001b) *Alkaliphilus transvaalensis* gen. nov., sp. nov., an extremely alkaliphilic bacterium isolated from a deep South African gold mine. *Int J Syst Evol Microbiol* 51:1245–1256
- Takai K, Gamo T, Tsunogai U, Nakayama N, Hirayama H, Nealson KH, Horikoshi K (2004) Geochemical and microbiological evidence for a hydrogen-based, hyperthermophilic subsurface lithoautotrophic microbial ecosystem (HyperSLiME) beneath an active deep-sea hydrothermal field. *Extremophiles* 8:269–282
- Takai K, Moyer CL, Miyazaki M, Nogi Y, Hirayama H, Nealson KH, Horikoshi K (2005) *Marinobacter alkaliphilus* sp. nov., a novel alkaliphilic bacterium isolated from seafloor alkaline serpentine mud from Ocean Drilling Program (ODP) Site 1200 at South Chamorro Seamount, Mariana Forearc. *Extremophiles* 9:17–27
- Takai K, Nakamura K, Toki T, Tsunogai T, Miyazaki M, Miyazaki J, Hirayama H, Nakagawa S, Nunoura T, Horikoshi K (2008a) Cell proliferation at 122 °C and isotopically heavy CH₄ production by a hyperthermophilic methanogen under high pressures cultivation. *Proc Natl Acad Sci U S A* 105:10949–10954

- Takai K, Nunoura T, Ishibashi J, Lupton J, Suzuki R, Hamasaki H, Ueno Y, Kawagucci S, Gamo T, Suzuki Y, Hirayama H, Horikoshi K (2008b) Variability in the microbial communities and hydrothermal fluid chemistry at the newly discovered Mariner hydrothermal field, southern Lau Basin. *J Geophys Res* 113:G02031
- Takai K, Mottl MJ, Nielsen SH, The Expedition 331 Scientists (2011) Proc IODP Exp 331. doi:<https://doi.org/10.2204/iodp.proc.331.2011> doi:<https://doi.org/10.2204/iodp.proc.331.2011>
- Takai K, Nakamura K, LaRowe D, Amend JP (2014) Chapter 2.4 – Life at seafloor extremes. In: Stein R, Blackman D, Inagaki F, Larsen HC (eds) Earth and life processes discovered from seafloor environments a decade of science achieved by the Integrated Ocean Drilling Program (IODP), *Dev Mar Geol* 7. Elsevier, Amsterdam, pp 149–174
- Takami H, Inoue A, Fuji F, Horikoshi K (1997) Microbial flora in the deepest sea mud of the Mariana Trench. *FEMS Microbiol Lett* 152:279–285
- Valentine DL (2007) Adaptations to energy stress dictate the ecology and evolution of the Archea. *Nat Rev Microbiol* 5:316–323
- Von Damm KL (1995) Controls on the chemistry and temporal variability of seafloor hydrothermal fluids. In: Humphris SE, Zierenberg RA, Mullineaux LS, Thomson RE (eds) Seafloor hydrothermal systems: physical, chemical, biological, and geological interactions, *Geophysical Monograph* 91. American Geophysical Union, Washington, DC, pp 222–247
- Vreeland RH, Rosenzweig WD, Powers DW (2000) Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature* 407:897–900
- Wilcock WSD, Fisher AT (2004) Geophysical constraints on the seafloor environment near Mid-Ocean Ridge. In: Wilcock WSD, Delong EF, Kelley DS, Baross JA, Cary SC (eds) The seafloor biosphere at mid-ocean ridges, *Geophysical Monograph* 144. American Geophysical Union, Washington, DC, pp 51–74
- Yanagawa K, Ijiri A, Breuker A, Sakai S, Miyoshi Y, Kawagucci S, Noguchi T, Hirai M, Schippers A, Ishibashi J, Takaki Y, Sunamura M, Urabe T, Nunoura T, Takai K (2017) Defining boundaries for the distribution of microbial communities beneath the sediment-buried, hydrothermally active seafloor. *ISME J* 11:529–542
- Yayanos AA (1986) Evolutional and ecological implications of the properties of deep-sea barophilic bacteria. *Proc Natl Acad Sci U S A* 83:9542–9546

Chapter 21

What Geology and Mineralogy Tell Us About Water on Mars



Tomohiro Usui

Abstract Since Mars has attracted much interest as a potentially accessible habitable planet, the greatest number of spacecraft has been sent to this planet among any of the other extraterrestrial bodies. The Mars exploration has provided evidence for a variety of water-related geological activities: fluvial landforms, paleo-oceans and lakes, and aqueous alteration and weathering of the surface materials. These geologic observations clearly indicated the existence of liquid water on the surface of Mars, while the most recent investigations have uncovered the possible existence of subsurface water (ice) world, which may be more favorable to extant or even present life on Mars.

Keywords Water · Subsurface · Mars exploration · Surface geology

21.1 Introduction

The surface geology and mineralogy record the unique history of planet Mars, which experienced dynamical transition from an ancient, water-rich Earth-like environment to the present-day cold and dry surface condition. Numbers of Mars exploration missions have provided compelling evidence for the presence of surface liquid water during the early geologic eras of Mars (pre-Noachian and Noachian: ~3.7 to 4.5 Ga, and Hesperian: ~3.0 to 3.7 Ga). Observations of widespread fluvial landforms, dense valley networks, evaporites, and hydrous minerals that are commonly formed by aqueous processes imply that Mars had an active hydrological cycle with lakes and possibly oceans (Carr 2006; Ehlmann and Edwards 2014). In contrast to the ancient watery environment, the surface of Mars is relatively cold and dry today. The recent desert-like surface conditions, however, do not necessarily indicate a lack of surface or near-surface water. Massive deposits of ground ice

T. Usui (✉)

Department of Solar System Sciences, Institute for Space and Astronautical Science,
Japan Aerospace Exploration Agency, Sagami-hara, Kanagawa, Japan
e-mail: usui.tomohiro@jaxa.jp

and/or icy sediments have been proposed based on subsurface radar sounder observations (Mouginot et al. 2012; Castaldo et al. 2017). Furthermore, landforms that appear to be glacial commonly occur even in the equatorial regions; this has led to the hypothesis that, like Earth, Mars had recent ice ages when water-ice would have globally covered the surface (Head et al. 2003, 2005). This chapter summarizes the surface geology and mineralogy and the evolution of water reservoirs on Mars, which is crucial to our understanding of the transition of climate and near-surface environments and habitability on Mars. The evolution of atmosphere and the potential for life on Mars are reviewed in Chaps. 22 and 23, respectively.

21.2 Review of Surface Geology and Mineralogy

A fundamental and noticeable surface feature of Mars is the global dichotomy in close relationship between topography and ages of geologic units. The global dichotomy divides the relatively old southern highlands from the young northern lowlands (Carr 2006) (Fig. 21.1). The southern highlands are heavily cratered ancient (Noachian) terrain, whereas the northern lowlands are smooth Hesperian-to-Amazonian-aged terrains covered with layered sedimentary and volcanic deposits. Other than the dichotomy, the largest positive topographic feature is the Tharsis bulge (~10 km high and 5000 km across). The volcanic surfaces around the bulge are relatively young (Amazonian), although Tharsis itself has been a locus of plume volcanism for billions of years (e.g., Vaucher et al. 2009).

Apparent water-related, fluvial surface landforms are mostly distributed in Noachian and Hesperian (>3 Ga) terrains. Outflow channels, characterized by wide-width (typically 10–100 km across), low sinuosity, and high width-depth ratios, are interpreted to be formed by huge catastrophic floods (Carr 2006). Their water sources are not fully understood but probably related to buried aquifers or rapid melting of surface or ground ice (Tanaka 1997). On the other hand, valley networks have high sinuosity, much narrow width (less than a few km), and length up to hundreds and even thousands of kilometers long, which is indicative of drainage of surface waters. The valley network activities were most intense in the late Noachian to early Hesperian (Fassett and Head 2008). Young Amazonian terrains still include local observations of fluvial landforms, including poorly dendritic valleys, channels, and recent gullies (Malin and Edgett 2000; Costard et al. 2002; Schon et al. 2009; Fassett et al. 2010; Kite et al. 2013a, b). Gullies are small and linear features incised into steep slopes. Their formation mechanism and the source of water are still under debate: e.g., seepage of groundwater (Malin and Edgett 2000) vs. melting of snow deposited on steep slope (Costard et al. 2002).

Another noticeable global topographic feature of water-related landforms is putative paleo-shorelines (Head et al. 1999) (Fig. 21.2). Topographic features of putative paleo-shorelines suggest that large bodies of standing water must have once occupied the northern lowlands. Shoreline-demarcation studies of the northern lowlands point to several contacts that yield variable sizes of paleo-oceans estimated to

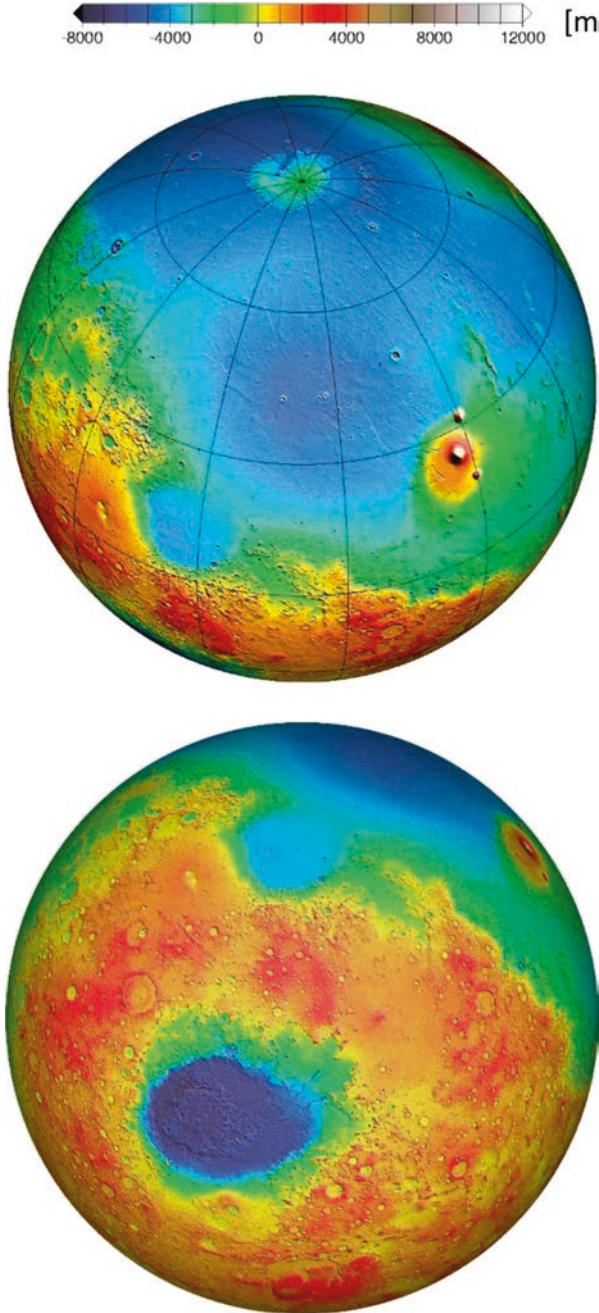


Fig. 21.1 Mars topography based on data from the Mars Orbiter Laser Altimeter (MOLA) on Mars Global Surveyor (Credit NASA/GFSC)

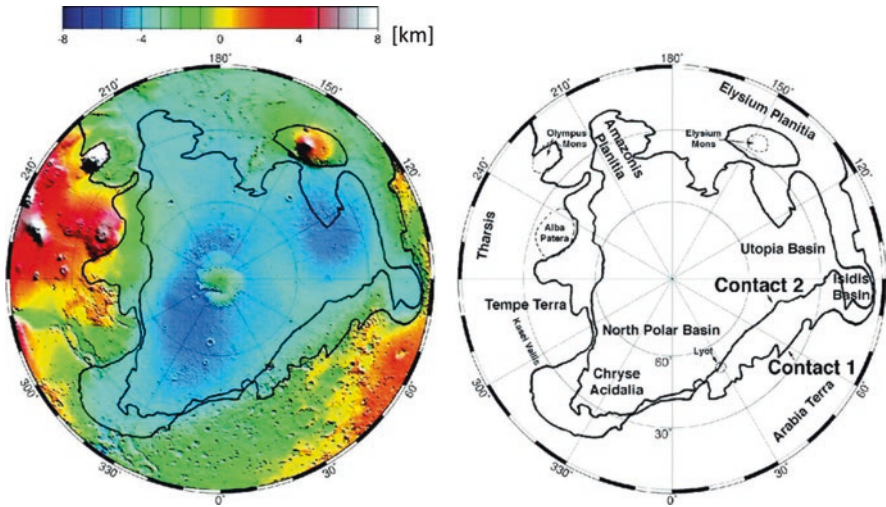


Fig. 21.2 Projection of MOLA topography from the North Pole (left). Black lines indicate positions of contacts (contacts 1 and 2), also shown on the right. (After Head et al. 1999)

range from $\sim 2 \times 10^7$ to 2×10^8 km³ (corresponding to global equivalent layers (GEL) of 130 m to 1500 m, respectively) (Kurokawa et al. 2014 and references therein). This variation has been interpreted to reflect the historical change in the ocean volume. For example, two major contacts (contact 1, Arabia shoreline, and contact 2, Deuteronilus shoreline) individually represent the larger Noachian and smaller Hesperian oceans, respectively (Parker et al. 1993; Clifford and Parker 2001; Carr and Head 2003; Di Achille and Hynes 2010). No contacts indicative of Amazonian oceans are reported. To conclude, the changes in the character of fluvial features and the size of paleo-ocean volumes clearly suggest the general trend of decline of liquid water activities on the surface from the ancient Noachian/Hesperian to the relatively recent Amazonian (Fig. 21.3).

Along with topographic features, the historic change of water-related activities is traced by the mineral records. The topographic features provide geophysical information of the water activities (e.g., volume and intensity of the floods), whereas the mineral records provide means to study the evolution of aqueous environments (e.g., fluid chemistry). For example, global mineralogical mapping on Mars suggests the secular desiccation and acidification of near-surface environment (Bibring et al. 2006; Ehlmann and Edwards 2014) (Fig. 21.4). The surface mineralogy of early Noachian terrains is characterized by clay minerals such as Fe/Mg smectites, suggesting water-rich and near-neutral fluid conditions. The occurrence and distribution of carbonates and sulfate mineralogy in late Noachian to early Hesperian suggest the transition of fluid chemistry to more acidic conditions in this period. The young Amazonian terrains are dominated by anhydrous ferric oxides such as hematite, suggestive of acidic and water-poor conditions.

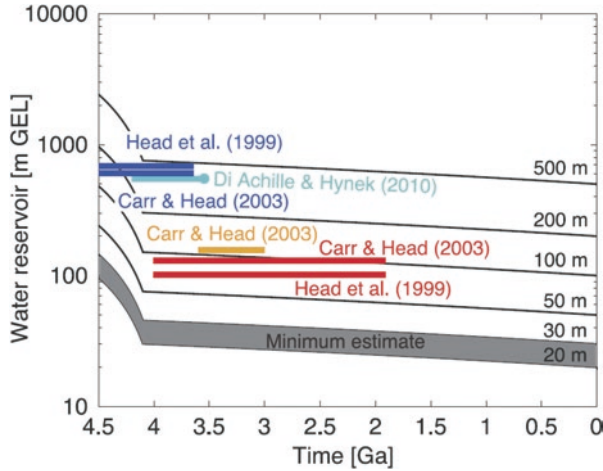


Fig. 21.3 Evolution of water reservoirs for different amounts of present water reservoirs (black lines) and geological estimates on the size of paleo-oceans (horizontal bar). The gray area indicates the evolution of surface water reservoir calculated based on a minimum water reservoir model. (After Kurokawa et al. 2014)

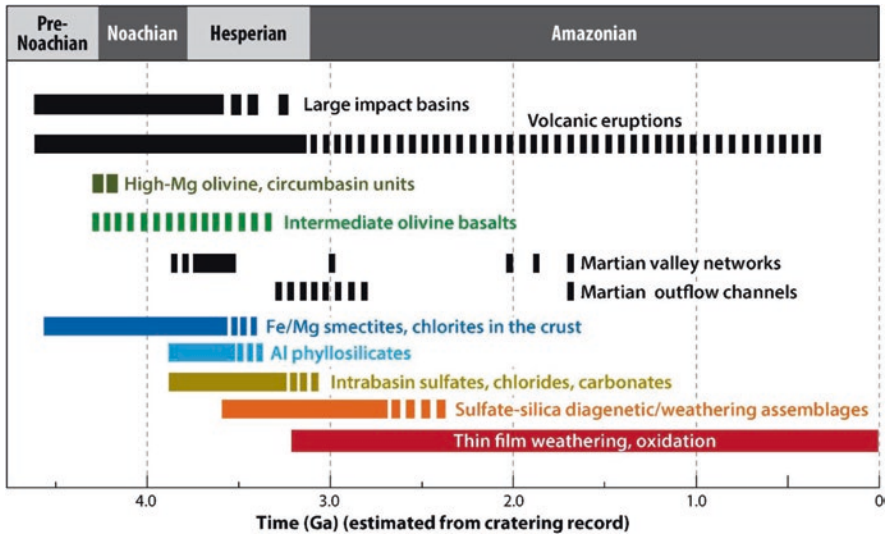


Fig. 21.4 Timeline of the major processes affecting the mineralogic composition of Mars and the ages of large-scale compositional units. (After Ehlmann and Edwards 2014)

21.3 Subsurface as a Potential Water Source and Refugium for Life on Mars

The global surface water inventory was originally estimated based on the size of putative paleo-oceans (Sect. 21.2). These geologic estimates are distinctly greater than the total water loss to space estimated based on atmospheric escape models (Kurokawa et al. 2014) (Chap. 22). The discrepancy between the geological and geophysical estimates of water volume hints at a “missing” water reservoir beneath the surface (Usui 2017). The widespread distribution of hydrated materials on the surface of Mars implies the existence of a crustal water reservoir, yet conventional spectroscopic observations are only able to see the surface veneer.

The thermodynamic modeling, together with remote sensing observations, has a means to provide the depth profile of hydrous materials and a reasonable estimate of the volume of a crustal water reservoir. Wade et al. (2017) examine the thermodynamic properties of water-bearing mafic crusts (basalt) and show that the Martian basalt can hold more structurally bound H₂O than terrestrial basalt and can effectively transport it to a greater depth (>90 km) within the Martian interior. They compute the stability of hydrous minerals in hydrated crusts and their bulk-rock densities in the pressure-temperature spaces along the geotherms of each planet. They conclude that over-plating and burial of hydrated crusts progressively hydrate the interior of Mars.

Ground ice is another candidate to account for the missing water reservoir on Mars. Subsurface radar sounder observations by Mars Express detected an electric anomaly (low dielectric constant) in the northern hemisphere, implying massive ice deposits that are interbedded with layers of sediment and volcanic materials at a depth of 60–80 m (Mouginot et al. 2012). The ground ice model is also proposed by hydrogen isotopic analysis of Martian meteorites (Usui et al. 2015) and crater morphology (Weiss and Head 2017). The crater morphological study indicates that the subsurface water-ice has a volume of $\sim 3 \times 10^7$ km³, which is comparable to the size of paleo-oceans. Furthermore, high-spatial resolution imaging by Mars Reconnaissance Orbiter detected the exposure of massive and layered ground ice in steep, pole-facing scarps created by erosion in mid-latitude deposits (Dundas et al. 2018). Near-infrared observations of the scarps indicate that the ground ice layers consist of pure H₂O (<1% soil) and each of them has a size of >100 m thick, extending downward from depths as shallow as ~ 1 to 2 m below the surface.

Due to the subfreezing temperature and low atmospheric pressure at the surface (Wordsworth 2016) (Chap. 22), strong UV radiation (Hassler et al. 2013), and the chaotic obliquity at least in the recent past (Laskar and Robutel 1993), the surface of Mars would have been a challenging habitat, while subsurface refugia for life may still exist (Cockell 2014). Moreover, the redox gradient between the oxidized surface and reduced subsurface, and availability of water, nutrients, and energy sources, might have become the subsurface more favorable to extant or even present life on Mars. Limitation of potential for life on Mars is further discussed in Chap. 23.

References

- Bibring JP et al (2006) Global mineralogical and aqueous mars history derived from OMEGA/ Mars express data. *Science* 312(5772):400–404
- Carr MH (2006) *The surface of Mars*. Cambridge University Press, Cambridge, UK
- Carr MH, Head JW (2003) Oceans on Mars: an assessment of the observational evidence and possible fate. *Jour Geophy Res* 108(E5):5042. <https://doi.org/10.1029/2002JE001963>
- Castaldo L et al (2017) Global permittivity mapping of the Martian surface from SHARAD. *Earth Planet Sci Lett* 462:55–65
- Clifford SM, Parker TJ (2001) The evolution of the Martian hydrosphere: implications for the fate of a primordial ocean and the current state of the northern plains. *Icarus* 154(1):40–79
- Cockell C (2014) The subsurface habitability of terrestrial rocky planets. In: *Microbial life of the deep biosphere, Life in extreme environments*, vol 1. De Gruyter, Berlin, pp 225–259
- Costard F et al (2002) Formation of recent Martian debris flows by melting of near-surface ground ice at high obliquity. *Science* 295(5552):110–113
- Di Achille G, Hynek BM (2010) Ancient ocean on Mars supported by global distribution of deltas and valleys. *Nat Geosci* 3(7):459–463
- Dundas CM et al (2018) Exposed subsurface ice sheets in the Martian mid-latitudes. *Science* 359(6372):199–201
- Ehlmann BL, Edwards CS (2014) Mineralogy of the Martian surface. *Annu Rev Earth Planet Sci Lett* 42(1):291–315
- Fassett CI, Head JW (2008) Valley network-fed, open-basin lakes on Mars: distribution and implications for Noachian surface and subsurface hydrology. *Icarus* 198(1):37–56
- Fassett CI et al (2010) Supraglacial and proglacial valleys on Amazonian Mars. *Icarus* 208(1):86–100
- Hassler DM et al (2013) Mars' surface radiation environment measured with the Mars science laboratory's curiosity rover. *Science* 343:1244797
- Head JW et al (1999) Possible ancient oceans on Mars: evidence from Mars Orbiter Laser Altimeter data. *Science* 286(5447):2134–2137
- Head JW et al (2003) Recent ice ages on Mars. *Nature* 426(6968):797–802
- Head J et al (2005) Tropical to mid-latitude snow and ice accumulation, flow and glaciation on Mars. *Nature* 434(7031):346–351
- Kite ES et al (2013a) Seasonal melting and the formation of sedimentary rocks on Mars, with predictions for the Gale Crater mound. *Icarus* 223(1):181–210
- Kite ES et al (2013b) Pacing early Mars river activity: embedded craters in the Aeolis Dorsa region imply river activity spanned \geq (1–20) Myr. *Icarus* 225(1):850–855
- Kurokawa H et al (2014) Evolution of water reservoirs on Mars: constraints from hydrogen isotopes in martian meteorites. *Earth Planet Sci Lett* 394:179–185
- Laskar J, Robutel P (1993) The chaotic obliquity of the planets. *Nature* 362:608–612
- Malin MC, Edgett KS (2000) Evidence for recent groundwater seepage and surface runoff on Mars. *Science* 288(5475):2330–2335
- Mouginot J et al (2012) Dielectric map of the Martian northern hemisphere and the nature of plain filling materials. *Geophys Res Lett* 39(2):L02202
- Parker TJ et al (1993) Coastal geomorphology of the Martian northern plains. *J Geophys Res Planet* 98(E6):11061–11078
- Schon SC et al (2009) Unique chronostratigraphic marker in depositional fan stratigraphy on Mars: evidence for ca. 1.25 Ma gully activity and surficial meltwater origin. *Geology* 37(3):207–210
- Tanaka KL (1997) Sedimentary history and mass flow structures of Chryse and Acidalia Planitiae, Mars. *J Geophys Res Planet* 102(E2):4131–4149
- Usui T (2017) Martian water stored underground. *Nature* 552:339–340
- Usui T et al (2015) Meteoritic evidence for a previously unrecognized hydrogen reservoir on Mars. *Earth Planet Sci Lett* 410:140–151

- Vaucher J et al (2009) The volcanic history of central Elysium Planitia: implications for martian magmatism. *Icarus* 204(2):418–442
- Wade J et al (2017) The divergent fates of primitive hydrospheric water on Earth and Mars. *Nature* 552(7685):391
- Weiss DK, Head JW (2017) Evidence for stabilization of the ice-cemented cryosphere in earlier Martian history: implications for the current abundance of groundwater at depth on Mars. *Icarus* 288:120–147
- Wordsworth RD (2016) The climate of early Mars. *Annu Rev Earth Planet Sci Lett* 44:381–408

Chapter 22

Atmosphere of Mars



Hiromu Nakagawa

Abstract It is believed that Mars underwent drastic climate change, changing its environment from warm and wet to cold and dry. This gives rise to the idea that Mars may have hosted life in the past and, indeed, may do so even today. Atmospheric evolution is thus an important key to understanding the history of Martian habitability. However, precise estimates of past atmospheric inventories including water, and their loss mechanisms, are difficult to be obtained. Recent studies have highlighted various interesting facts related to (i) the efficiency of mass transport from the lower to upper atmospheric reservoir and (ii) the deep energetic particle precipitation into the atmosphere from space. These new insights tell us that Mars is a mutually coupled system comprising the planet's surface, lower and upper atmospheres, and the surrounding space environment. These relationships potentially imply an upward revision of the estimate of total atmospheric loss to space. Another relevant issue relates to the indirect signs of life in the Martian atmosphere. Scientists are particularly intrigued by clear evidence of a biological/geological signature, such as methane (CH_4) in the Martian atmosphere. Although the presence of CH_4 is still under debate because of large measurement uncertainties, the forthcoming ESA-Roscosmos mission, which employs the Trace Gas Orbiter (TGO), will settle questions on the existence of this gas and its origin.

Keywords Mars · Water · Volatile · Habitability · Methane

22.1 Introduction

The atmospheric environment of current-day Mars is far from habitable. The air pressure, 6 mbar, is less than 1% of that on Earth. Although the atmosphere is mostly composed of carbon dioxide (CO_2) (~96%), it creates a weak greenhouse effect. The temperature is usually well below the freezing point of water, with an overall average of about $-58\text{ }^\circ\text{C}$ ($215\text{ }^\circ\text{K}$). The surface and atmosphere are heavily affected

H. Nakagawa (✉)
Tohoku University, Sendai-shi, Miyagi, Japan
e-mail: hnakagawa@tohoku.ac.jp

by energetic photons and particles that easily penetrated it owing to insufficient magnetospheric and atmospheric shielding. Although Mars is dry and frozen today, the geological evidence points to drastic climate change in Martian history. Mars had at least some warm and wet durations in the past, enough to allow for liquid water to be stable (Head et al. 1999; Parker et al. 1993).

Mars must somehow have lost most of its atmosphere and water. The evolution of the Mars environment is thought to have proceeded by some combination of atmospheric escape to space and surface adsorption. Especially, a lack of global magnetic field on Mars must have a significant impact on the atmospheric escape. The weaker magnetosphere would have allowed solar winds to strip away much of its atmosphere to space. Consequently, the planet's thinner atmosphere would reduce greenhouse warming. Although previous studies have proposed a variety of processes to explain the atmospheric escape of Mars to space (Shizgal and Arkos 1996; Chassefière and Leblanc 2004; Lundin et al. 2009), its understanding is shrouded in mystery owing to lack of measurements and difficulty in validation by theoretical studies (Dubinin et al. 2011; Lundin 2011; Harnett and Winglee 2006; Ma et al. 2004). Importantly, since the precise mechanisms in which Mars lost its atmosphere and water have not yet been settled, precise estimations of the atmospheric and water amounts that have escaped to space in the past remain elusive.

The upper atmosphere is the primary reservoir for atmospheric escape. Owing to the technical difficulties associated with observations, datasets of the upper atmosphere are limited. However, the last decade has seen a renewed interest in the upper atmosphere, thanks to the aggressive explorations by the Mars Reconnaissance Orbiter (MRO), Mars Express (MEX), and Mars Volatile Evolution (MAVEN) missions. In this chapter, we briefly introduce recent findings relevant to the atmospheric evolution of Mars.

22.2 Atmospheric Loss to Space

Jakosky et al. (2017) estimated the amount of gas lost to space through time using measurements by MAVEN. Fractionation of argon isotopes in the upper atmosphere occurs as a result of loss of gas to space by pickup ion sputtering. Sputtering, which occurs when neutralized solar wind ions impart their large energies to surrounding particles by collisions, may have been important on early Mars, after it lost its magnetic field and was no longer shielded from the solar wind. They estimated the degree of fractionation of $^{38}\text{Ar}/^{36}\text{Ar}$ between the homopause (~120 km) and the exobase (~200 to 300 km) altitudes, where the molecular diffusion dominates and mixing ratios of lighter species increase with altitude, to determine the fraction of the total argon removed from Mars (Fig. 22.1). The measurements indicate that 66% of the argon was lost to space during the past 4 Gyr. They can use the result to estimate the amount of other gases that would have been lost by the same sputtering mechanism, such as oxygen. Oxygen originates from either CO_2 or H_2O

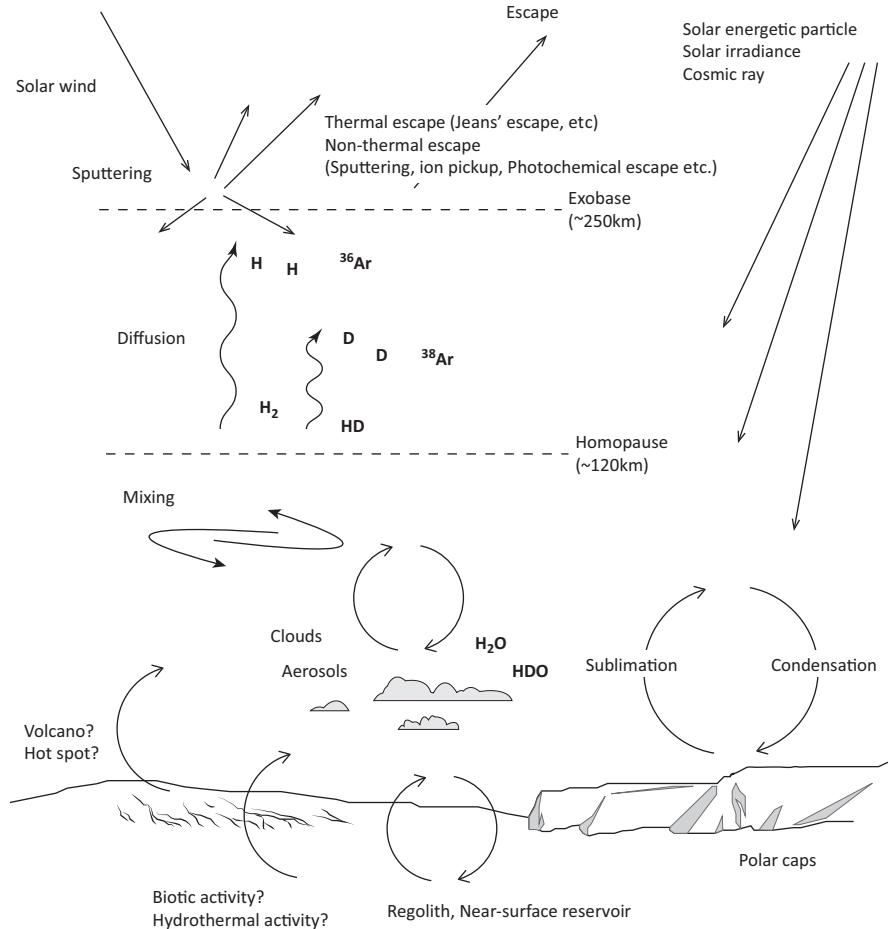


Fig. 22.1 The atmospheric evolution on Mars. The atmosphere is affected by energetic photons and particles from space. The sublimation-condensation process between surface and atmosphere depletes heavy water in the gas phase. The heavy isotope enrichment in the upper atmosphere occurred by the loss of light isotopes to space. The fractionation rate in the upper atmosphere is thus crucial to estimate the atmospheric loss to space

via photodissociation process by sunlight. Since the lost oxygen would be significant as predicted in the model (Luhmann et al. 1992), their result proposed that the sputtering loss of CO₂ can approach a bar or more if we assume the lost oxygen originates primarily from sputtering of CO₂. The loss rate might be much higher in the early stage of Mars' history because of the more intense solar extreme UV radiation and solar wind activities. CO₂ also can be removed by other processes, including pickup by the solar wind and photochemically driven escape (Chassefière et al. 2007). Hu et al. (2015) attempted to analyze the enrichment of ¹³C in the Martian atmosphere by considering photochemical escape and exchange between

carbon reservoirs. Their results indicated that the early CO₂ atmosphere might have been characterized by pressures of up to 1.8 bar, about 1.8 times higher pressure than current Earth surface. A large fraction of Mars' atmospheric gas is suggested to have been lost, contributing to drastic climate change on the planet's surface. More realistic estimates of current and ancient fractionation rates, including other escape processes or other species, will be investigated by the MAVEN mission (Lillis et al. 2015; Leblanc et al. 2015; Jakosky et al. 2018).

On the other hand, it is well known that CO₂ cannot warm early Mars because CO₂ condenses and reflects a large fraction of solar energy back to space (Kasting 1991). Pollack et al. (1987) was the first to argue for a warmer and wetter early Mars using climate models. Forget et al. (2013) reported the numerical simulations contending that a CO₂ atmosphere of up to 7 bars could not have raised the annual mean temperature above 0 °C anywhere on early Mars. On the other hand, Ramirez et al. (2014) and Wordsworth et al. (2017) noted that the effects of H₂ and CH₄ on the greenhouse warming in addition to the CO₂ and H₂O could raise annual mean surface temperatures above the freezing point of water. Wordsworth et al. (2015) examined two scenarios, “warm and wet” and “cold and icy,” in order to reconcile the early Martian hydrological cycle and geological features on the surface, and argued that the geological evidence may favor the “cold and icy” scenarios in their model. Seminal reviews of these two scenarios can be found in Craddock and Howard (2002) and Wordsworth et al. (2013). The “faint young Sun” paradox on Mars might be more severe than that in the Earth's case due to the greater distance of the former from the Sun. These quantitative studies will be extended to solve the question of whether there were ever long-term conditions that could have allowed a surface biosphere to flourish on Mars. Future human explorations may constrain the ancient atmospheric pressure if, for example, raindrop imprints in tuffs were to be found (e.g., Som et al. 2012).

22.3 Water Loss to Space

How much water might Mars have lost during the history since its birth? One way to estimate this amount is by measuring the deuterium-to-hydrogen ratio (D/H) in the Martian atmosphere, because D is heavier than ordinary H and therefore escapes to space less easily (Fig. 22.1). The D/H ratio in atmospheric water vapor on Mars has been deduced from remote spectroscopy as being five to six times higher than the corresponding global average value in Earth's oceans (Vienna Standard Mean Ocean Water, VSMOW, HDO/H₂O = 3.11 × 10⁻⁴) in a global average (Owen 1992; Owen et al. 1988; Krasnopolsky 2000), which is consistent with the findings of Mars Science Laboratory (MSL), Curiosity, and in situ measurements of 6 ± 1 obtained at Gale Crater (Webster et al. 2013). An equation for the Rayleigh fractionation of H between an initial D/H, (D/H)_{initial}, and a current D/H, (D/H)_{later}, can be written as below (see Donahue 1995; Krasnopolsky et al. 1998):

$$(M_{\text{initial}} / M_{\text{later}})^{1-f} = \left[(D/H)_{\text{later}} / (D/H)_{\text{initial}} \right] \quad (22.1)$$

$$f = 2(\varphi_D / \varphi_H) / [\text{HDO} / \text{H}_2\text{O}] \quad (22.2)$$

Here, M_{initial} is the initial hydrogen reservoir mass, which is assumed to be water, M_{later} is the current water reservoir, f is a fractionation factor that indicates the relative efficiency of D escape and H escape, and φ_D and φ_H are the escape rates of D and H, respectively. Applying $(D/H)_{\text{initial}} = 1.28$ VSMOW, $(D/H)_{\text{later}} = 5.5$ VSMOW, $M_{\text{later}} = 25$ m of water, and $f = 0.4$, Catling and Kasting (2017) get $M_{\text{initial}} = 284$ m as the global equivalent layer (GEL) of water.

Recent high spatial- and spectral-resolution remote spectroscopy data from the ground-based facilities suggests notably higher deuterium enrichment than was found in previous globally averaged observations (Aoki et al. 2015a; Encrenaz et al. 2015; Villanueva et al. 2015). This difference is explained by the full-disk measurements, which reflect the mean of diverse regions with high and low D/H , as revealed by their resolved maps. Although the model predicts a D/H variability of 15% at latitudes due to the condensation-evaporation processes that deplete heavy water in the gas phase (Montmessin et al. 2005), the strong local anisotropies, in the range between 1 and 10 VSMOW, require a more realistic model to account for several climatological processes acting on the isotopologues, and the results imply unknown or multiple reservoirs. Using $f = 0.02$, Villanueva et al. (2015) reported a value of 8 VSMOW for permanent polar caps, which exceeds 137 m GEL. Thus, the estimations by these equations, however, strongly depend on the fractionation factor, f . Chaffin et al. (2014) reported that φ_H could vary considerably, from 5 to 500 ($10^7 \text{ cm}^{-2} \text{ s}^{-1}$) over months or a shorter time period. Even $(D/H)_{\text{later}}$ could vary according to the recent measurements of H and D corona (Clarke et al. 2017). It is also noted that the solar activity in early phase could be so intense that the fractionation factor between D and H would have been small, implying that these water inventory estimates might be lower limits. More precise estimates thus require spatially and temporally resolved measurements of hydrogen isotopes.

22.4 Deep Energetic Particle Precipitation into the Atmosphere from Space

The Spectroscopy for Investigation of Characteristics of the Atmosphere of Mars (SPICAM) UV spectrograph onboard MEX first discovered auroral emissions on Mars (Bertaux et al. 2005). This “discrete” aurora identified on Mars is a highly concentrated and localized emission around magnetic field anomalies in the southern hemisphere (Fig. 22.2) (Leblanc et al. 2006). The intensities range from 100 to 2000 R (Rayleigh = $10^6/4\pi$ photons/cm²/str/s) for CO Cameron bands (180–260 nm) and from 10 to 160 R for the CO₂⁺ doublet (288 nm). The altitude of the emissions is ~130 km. On Earth, familiar aurora occurs near the footprint of the dipole

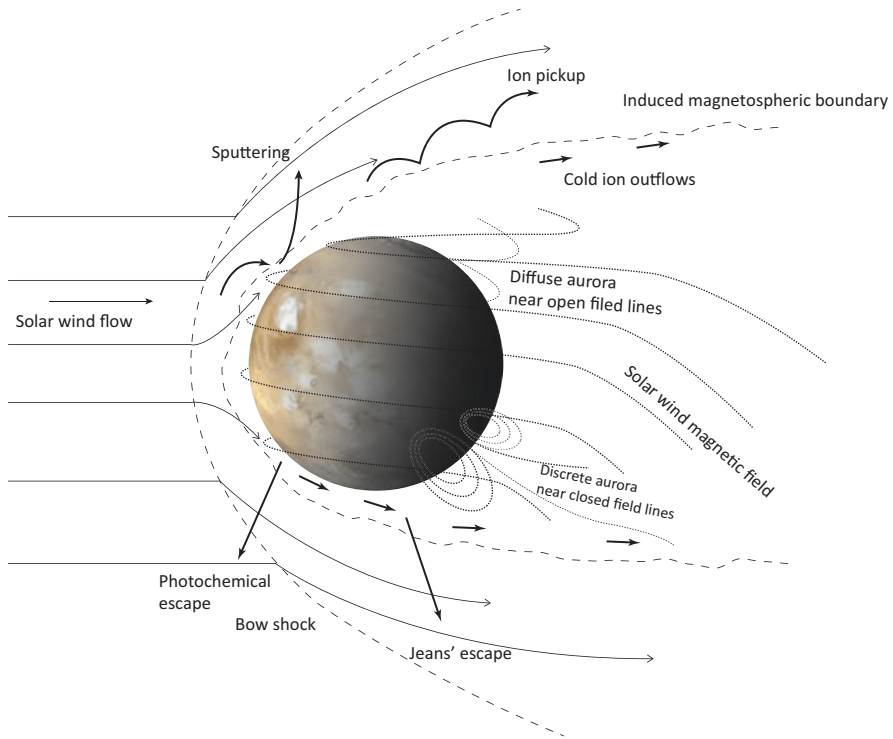


Fig. 22.2 A schematic illustration of Mars space environment. The weaker magnetosphere would have allowed solar winds to strip away much of its atmosphere to space. The magnetic fields of solar winds drape over the Mars. The discrete auroras are localized around magnetic field anomalies in the southern hemisphere. The diffuse auroras, which indicate deep energetic particle penetration, span across the northern hemisphere

magnetic field, where interactions with the solar wind electric field can cause reconnection and energetic particles within the magnetosphere. This discovery confirmed this scenario could be applied on Mars' patchy magnetic field. The auroral radiation is emitted by atmospheric constituents that are excited by precipitating energetic particles (electrons and ions). The light from an aurora is proportional to the deposition of energy into the atmosphere by the primary particles. The height distribution of aurora is related to the energy of the precipitating particles as well as the atmospheric composition. Since their precipitations heat up, expand the upper atmosphere, and also contribute the ionization process of the atmosphere, the measurement of aurora emission is key to understand the deposition of energy into the atmosphere from the space. The behavior of Martian discrete aurora might provide another aspect of auroral effect on the atmospheric evolution under the different condition of intrinsic magnetic field of planet.

The Imaging Ultraviolet Spectrograph (IUVS) spectroscopy onboard MAVEN detected another type of aurora, suggesting more widespread occurrences with

increased solar activity (Schneider et al. 2015). Schneider et al. (2015) reported the low-altitude, “diffuse” auroras spanning across Mars’ northern hemisphere, coincident with a solar energetic particle (SEP) outburst. The emission extended down to an altitude of ~60 km. The intensities reached up to several hundreds R at the CO_2^+ doublet. Deep precipitation in the upper atmosphere requires extremely energetic electron fluxes up to 100 keV. This is consistent with the measurements of SEP by MAVEN instrument. Diffuse auroras may have additional effects on atmospheric processes. Incident particles ionize and dissociate atmospheric species deeply, as well as heat the target atmosphere. These effects may lead to increased atmospheric escape rates: ionized particles at sufficiently high altitudes can escape via outflow processes, and atmospheric heating can lead to increased thermal escape. IUVS results thus offer a new paradigm for solar/stellar influences on non-magnetized planets. The atmospheric effects of auroral energy input are not known well on Mars. Diffuse auroras provide visible tracers for particle penetration, allowing a better understanding of Mars’ interaction with the space environment. More energy is deposited deep in the atmosphere than previously known.

22.5 Energy and Mass Transport into the Upper Atmosphere from Below

The highly variable nature of exospheric H escape has been identified (Chaffin et al. 2014; Clarke et al. 2014). Because the production timescale of H_2 (major source of exospheric H to escape) from H_2O in the lower atmosphere is predicted to be a $\sim 10^4$ to 10^5 years by the model (Hunter and McElroy 1970), the H escape rate did not vary from month-to-month or from year-to-year. In contrast, Chaffin et al. (2014) and Clarke et al. (2014) demonstrated that altitude profiles of Lyman-alpha sunlight scattered from exospheric H imply an order-of-magnitude decline in the H escape rate in a timescale of a season or shorter. This may represent relaxation from a temporarily enhanced escape rate, possibly due to a global dust storm in which H_2O was carried from the surface to the upper atmosphere, which dramatically increases H production. These results show that H escape is more variable than previously suspected, potentially resulting in much larger integrated water loss over Martian history.

It is interesting to note that high-altitude (above 60 km) H_2O was first identified by SPICAM occultations (Maltagliati et al. 2011, 2013) especially in the southern summer, which happens to be a dusty season (at a solar longitude of 240° or later). Chaffin et al. (2017) demonstrated that the high-altitude H_2O can considerably increase H escape on the weekly timescale, which can potentially explain the observations mentioned above. Maltagliati et al. (2013) also showed the links between such high-altitude H_2O and aerosols in their vertical profiles within a short timescale. This implies the importance of aerosols for key processes in the Martian water cycle and climate as a whole. Importantly, the new pathway of water loss

proposed by these studies implies much higher loss to space, in addition to the diffuse-limited escape of H_2 (Catling and Kasting 2017).

One of the important findings of the MEX mission is the significant amount of heavy molecular ion escape from Mars. Carlsson et al. (2006) showed that the escape flux of O_2^+ is comparable to that of O^+ , and CO_2^+ flux is about one-fifth of these values at the exobase at around 250 km. The large amount of escape is unexpected as per current theories, since the heavy molecular ions such as CO_2^+ have smaller scale heights than the lighter O^+ , and it is difficult to pull out these species from the bottom of the upper atmosphere, unless some mechanism exists to transport CO_2 molecules to high altitudes in the thermosphere, to provide a source of CO_2^+ ions. In order to understand the mechanism, it is essential to determine how the atmospheric escape is driven by the energy and momentum transfer from the solar wind from above and upward mass transport from below.

One possible scenario for upward transport from the lower atmosphere is the enhanced diffusion caused by gravity waves (GWs) of lower atmospheric origin. GWs have significant effects on large-scale winds (Medvedev et al. 2011), thermal balance (Medvedev and Yiğit 2012), and density (Medvedev et al. 2016) in the upper atmosphere. The vertical mixing must influence the homopause height (Imamura et al. 2016). The location of the homopause influences the upper atmosphere composition, thereby allowing the species to escape to space. In situ measurements of the upper atmosphere by MAVEN inferred the substantial variations of the homopause and exobase altitudes (Jakosky et al. 2017). They imply that both vary largely as a result of the behavior of the lower atmosphere. MAVEN data also revealed that the atmospheric waves exist ubiquitously in the upper atmosphere (Bougher et al. 2015; England et al. 2017; Terada et al. 2017). The average amplitude of GWs in the Martian upper thermosphere is $\sim 10\%$ on the dayside and $\sim 20\%$ on the nightside, which is about two and ten times larger than those on Venus and the low-latitude region of Earth, respectively (Kasprzak et al. 1988; Bruinsma and Forbes 2008). Answering questions about the upper atmospheric sources of these waves and their possible links with those in the troposphere is a key to understanding the efficient upward transport process from below.

22.6 Methane in the Atmosphere

Traces of methane (CH_4), averaging ~ 10 part per billion in volume (ppbv, 10^{-9}), were first reported, based on the ground-based facilities from Earth and MEX in 2003–2004 (Mumma et al. 2003; Formisano et al. 2004; Krasnopolsky et al. 2004). The potential significance of CH_4 is that it could originate from geological (e.g., serpentinization) or even biological underground sources (e.g., subsurface microorganisms) (Atreya et al. 2007). The following studies forwarded arguments regarding high variability along both time and space. The Planetary Fourier Spectrometer (PFS) measurements onboard MEX showed CH_4 enhancement of ~ 60 ppbv over the north polar cap during the summer season (Geminali et al. 2011). Font and

Marzo (2010) suggested different spatial and seasonal distributions of CH₄ by using the Thermal Emission Spectrometer (TES) onboard the Mars Global Surveyor (MGS), with peak abundance ~70 ppbv over Tharsis, Arabia Terra, and Elysium. Using high-resolution infrared spectroscopy on ground-based facilities, Mumma et al. (2009) found extended plumes of CH₄ (~40 ppbv) during the northern summer above Terra Sabae, Nili Fossae, and Syrtis Major in 2003. However, they reported no detection of CH₄ after 2006 (Villanueva et al. 2013), the upper limit being 7 ppbv. In contrast, Krasnopolsky (2012) claimed that CH₄ of 0–20 ppbv was detected over Valles Marineris using ground-based observations in 2006.

These reports that the CH₄ is variable, both spatially and on timescales of days to months, are difficult to reconcile with the photochemical stability of CH₄, which had a residence time of ~300 years and be well mixed by atmospheric circulations (Lefèvre and Forget 2009). A key question is the source of the CH₄ release. The current level of geophysical activity, such as volcanism and outgassing, has not been detected so far (Christensen et al. 2003; Krasnopolsky 2012; Villanueva et al. 2013; Khayat et al. 2015). On the other hand, observations of hydrogen peroxide (H₂O₂), which is an important factor in the oxidizing capacity of the Martian atmosphere, found a range from 0 to 50 ppbv (Encrenaz et al. 2008; Aoki et al. 2015b). This suggests the presence of strong CH₄ sinks not subject to atmospheric oxidation. Furthermore, owing to telluric or Martian contaminations, the CH₄ signals showed potentially a large uncertainty or ambiguity for reproducing the other spectral features surrounding the target line (Zahnle 2015).

Curiosity has detected measurable and variable levels of CH₄ gas along its travels through the Gale Crater. The first observation reported by the Sample Analysis at Mars (SAM) was a non-detection of CH₄ (Webster et al. 2013). The latest results of Curiosity then show values of CH₄ of 0.69 ± 0.25 ppbv and elevated levels of 7.19 ± 2.06 ppbv (Webster et al. 2014). There is still some controversy as to whether it has actually detected CH₄ from Mars or is only detecting CH₄ that the spacecraft brought with it from Earth, but follow-up and further refinement of the existing CH₄ measurements with Curiosity and coming landers will improve our understanding of the presence or absence of biological processes. The forthcoming orbiter exploration by Trace Gas Orbiter (TGO) will perform a sensitive search for CH₄, using high-resolution and highly sensitive solar occultations (the upper limit of several tens parts per trillion in volume (pptv, 10⁻¹²), is state of the art (Vandaele et al. 2015; Korabev et al. 2015). The determination of the exact origin requires measurements of CH₄ isotopologues (¹³CH₄, CH₃D) and of other trace gases related to possible CH₄ production/sink processes.

22.7 Conclusion

Recent findings have refined our understanding of Mars; we now appreciate that it is a considerably mutually coupled system of the surface, lower and upper atmospheres, and the space environment than previously expected. All these

observations unveil a complex and dynamically rich atmosphere, strongly affected by the interactions from both above and below. In order to achieve a complete understanding of the upper atmosphere or the primary mechanism and location of atmospheric escape, it is crucial to consider all atmospheric regions, the Martian surface, near-Mars space, and their interactions. Recent results have shed considerable light on the potential amounts of the atmosphere and water that has escaped to space in the past. They suggest the importance of isotopes and minor trace species measurements in the Mars atmosphere, in that both are spatially and temporally resolved, leading to more accurate estimates for atmospheric and water losses. Among the trace gas species, methane is of importance, because of the relation to the past and extent life. The spatial and temporal distribution of methane will be revealed by the ongoing mission of Trace Gas Orbiter.

References

- Aoki S, Giuranna M, Kasaba Y, Nakagawa H, Sindoni G, Geminale A, Formisano V (2015a) Search for hydrogen peroxide in the Martian atmosphere by the Planetary Fourier Spectrometer onboard Mars express. *Icarus* 245:177–183
- Aoki S, Nakagawa H, Sagawa H, Giuranna M, Sindoni G, Aronica A, Kasaba Y (2015b) Seasonal variation of the HDO/H₂O ratio in the atmosphere of Mars at the middle of northern spring and beginning of northern summer. *Icarus* 260:7–22
- Atreya SK, Mahaffy PR, Wong AS (2007) Methane and related trace species on Mars: origin, loss, implications for life, and habitability. *Planet Space Sci* 55:358–369
- Bertaux JL, Leblanc F, Witasse O, Quemerais E, Lilensten J, Stern SA, Sandel B, Korabiev O (2005) Discovery of an aurora on Mars. *Nature* 435:790. <https://doi.org/10.1038/nature03603>
- Bougher S, Jakosky B, Halekas J, Grebowsky J, Luhmann J, Mahaffy P, Connerney J, Eparvier F, Ergun R, Larson D, McFadden J, Mitchell D, Shneider N, Zurek R, Mazelle C, Andersson L, Andrews D, Baird D, Baker DN, Bell JM, Benna M, Brain D, Chaffin M, Chamberlin P, Chaugray JY, Clarke J, Collinson G, Combi M, Crary F, Cravens T, Crismani M, Curry S, Curtis D, Deighan J, Delory G, Dewey R, DiBraccio G, Dong C, Dong Y, Dunn P, Elrod M, England S, Eriksson A, Espley J, Evans S, Fang X, Fillingim M, Fortier K, Fowler CM, Fox J, Groller H, Guzewich S, Hara T, Harada Y, Holsclaw G, Jain SK, Jolitz R, Leblanc F, Lee CO, Lee Y, Lefèvre F, Lillis R, Livi R, Lo D, Ma Y, Mayyasi M, McClintock W, McEnulty T, Modolo Montmessin RF, Morooka M, Nagy A, Olsen K, Peterson W, Rahmati A, Ruhunusiri S, Russell CT, Sakai S, Sauvaud JA, Seki K, Steckiewicz M, Stevens M, Stewart AIF, Stiepen A, Stone S, Tennishev V, Thiemann E, Tolson R, Toublanc D, Vogt M, Weber T, Withers P, Woods T, Yelle R (2015) Early MAVEN Deep Dip campaign reveals thermosphere and ionosphere variability. *Science* 350:aad0459–aad0451
- Bruinsma SL, Forbes JM (2008) Medium- to large-scale density variability as observed by CHAMP. *Space Weather* 6:S08002. <https://doi.org/10.1029/2008SW0004111>
- Carlsson E, Fedorov A, Barabash S, Budnik E, Grigoriev A, Gunell H, Nilsson H, Sauvaud JA, Lundin R, Futaana Y, Holmström M, Andersson H, Yamauchi M, Winningham JD, Frahm RA, Sharber JR, Scherrer J, Coates AJ, Linder DR, Kataria DO, Kallio E, Koskinen H, Säles T, Riihelä P, Schmidt W, Kozyra J, Luhmann J, Roelof E, Williams D, Livi S, Curtis CC, Hsieh KC, Sandel BR, Grande M, Carter M, Thocaven JJ, McKenna-Lawler S, Orsini S, Cerulli-Irelli R, Maggi M, Wurz P, Bochsler P, Krupp N, Woch J, Fränz M, Asamura K, Dierker C (2006) Mass composition of the escaping plasma at Mars. *Icarus* 182:320–328

- Catling DC, Kasting JF (2017) Atmospheric evolution on inhabited and lifeless worlds. Cambridge University Press, UK
- Chaffin MS, Chafraï JY, Stewart I, Montmessin F, Schneider NM, Bertaux JL (2014) Unexpected variability of Martian hydrogen escape. *Geophys Res Lett* 41:314–320. <https://doi.org/10.1002/2013GL058578>
- Chaffin MS, Deighan J, Schneider NM, Stewart AIF (2017) Elevated atmospheric escape of atomic hydrogen from Mars induced by high-altitude water. *Nat Geosci* 10(3):174–178
- Chassefière E, Leblanc F (2004) Mars atmospheric escape and evolution; interaction with the solar wind. *Planet Space Sci* 52:1039–1058
- Chassefière E, Leblanc F, Langlais B (2007) The combined effects of escape and magnetic field histories at Mars. *Planet Space Sci* 55:343–357
- Christensen PR, Bandfield JL, Bell III JF, Gorelick N, Hamilton VE, Ivanov A, Jakosky BM, Kieffer HH, Lane MD, Malin MC, McConnochie T, McEwen AS, McSween HY Jr, Mehall GL, Moersch JE, Neason KH, Rice JW Jr, Richardson MI, Ruff SW, Smith MD, Titus TN, Wyatt MB (2003) Morphology and composition of the surface of Mars: Mars Odyssey THEMIS results. *Science* 300:2056
- Clarke JT, Bertaux JL, Chaufray JY, Gladstone GR, Quemerais E, Wilson JK, Bhattacharyya D (2014) A rapid decrease of the hydrogen corona of Mars. *Geophys Res Lett* 41:8013–8020. <https://doi.org/10.1002/2014GL061803>
- Clarke JT, Mayyasi M, Bhattacharyya D, Schneider NM, McClintock WE, Deighan JI, Stewart AIF, Chaufray JY, Chaffin MS, Jain SK, Stiepen A, Crismani M, Holsclaw GM, Montmessin F, Jakosky BM (2017) Variability of D and H in the Martian upper atmosphere observed with the MAVEN IUVS echelle channel. *J Geophys Res* 122:2336–2344. <https://doi.org/10.1002/2016JA23479>
- Craddock RA, Howard AD (2002) The case for rainfall on a warm, wet early Mars. *J Geophys Res* 107(E11):5111. <https://doi.org/10.1029/2001JE001505>
- Donahue TM (1995) Evolution of water reservoirs on Mars from D/H ratios in the atmosphere and crust. *Nature* 374:432–434. <https://doi.org/10.1038/374432a0>
- Dubinin E, Fraenz M, Fedorov A, Lundin R, Edberg N, Duru F, Vaisberg O (2011) Ion energization and escape on Mars and Venus. *Space Sci Rev* 162:173–211. <https://doi.org/10.1007/s11214-9831-7>
- Encrenaz T, Greathouse TK, Richter MJ, Bézard B, Fouchet T, Lefèvre F, Montmessin F, Forget F, Lebonnois S, Atreya SK (2008) Simultaneous mapping of H₂O and H₂O₂ on Mars from infrared high-resolution imaging spectroscopy. *Icarus* 195:547–556
- Encrenaz T, Greathouse TK, Lefèvre F, Montmessin F, Forget F, Fouchet T, DeWitt C, Richter MJ, Lacy JH, Bézard B, Atreya SK (2015) Seasonal variations of hydrogen peroxide and water vapor on Mars: further indications of heterogeneous chemistry. *Astron Astrophys* 578:A127. <https://doi.org/10.1051/0004-6361/201425448>
- England SL, Liu G, Yiğit E, Mahaffy PR, Elrod M, Benna M, Nakagawa H, Terada N, Jakosky B (2017) MAVEN NGISM observations of atmospheric gravity waves in the Martian thermosphere. *J Geophys Res* 122:2310–2335. <https://doi.org/10.1002/2016JA023475>
- Font S, Marzo GA (2010) Mapping the methane on Mars. *Astron Astrophys*:A51. <https://doi.org/10.1051/0004-6361/200913178>
- Forget F, Wordsworth R, Millour E, Madeleine JB, Kerber L, Leconte J, Marcq E, Haberle RM (2013) 3D modeling of the early martian climate under a denser CO₂ atmosphere: temperatures and CO₂ ice clouds. *Icarus* 222:81–99
- Formisano V, Atreya S, Encrenaz T, Ignatiev N, Giuranna M (2004) Detection of methane in the atmosphere of Mars. *Science* 306:1758. <https://doi.org/10.1126/science.11101732>
- Geminale A, Formisano V, Sindoni G (2011) Mapping methane in Martian atmosphere with PFS-MEX data. *Planet Space Sci* 59:137–148
- Harnett EM, Winglee RM (2006) Three-dimensional multifluid simulations of ionospheric loss at Mars from nominal solar wind conditions to magnetic cloud events. *J Geophys Res* 111:A09213. <https://doi.org/10.1029/2006JA011724>

- Head JW, Hiesinger H, Ivanov MA, Kreslavsky MA, Pratt S, Thomson BJ (1999) Possible ancient oceans on Mars: evidence from Mars orbiter laser altimeter data. *Science* 286:2134–2137. <https://doi.org/10.1126/science.286.5447.2134>
- Hu R, Kass DM, Ehlmann BL, Yung YL (2015) Tracing the fate of carbon and the atmospheric evolution of Mars. *Nat Commun*. <https://doi.org/10.1038/ncomms10003>
- Hunter DM, McElroy MB (1970) Production and escape of hydrogen on Mars. *J Geophys Res* 75(31):5989
- Imamura T, Watanabe A, Maejima Y (2016) Convective generation and vertical propagation of fast gravity waves on Mars: one- and two-dimensional modeling. *Icarus* 267:51–63
- Jakosky BM, Slipski M, Benna M, Mahaffy P, Elrod M, Yelle R, Stone S, Alsaeed N (2017) Mars' atmospheric history derived from upper-atmosphere measurements of $^{38}\text{Ar}/^{36}\text{Ar}$. *Science* 355:1408–1410
- Jakosky BM, Brain D, Chaffin M, Curry S, Deighan J, Grebowsky J, Halekas J, Leblanc F, Lillis R, Luhmann JG, Andersson L, Andre N, Andrews D, Baird D, Baker D, Bell J, Benna M, Bhattacharyya D, Bougher S, Bowers C, Chamberlin P, Chaufray J-Y, Clarke J, Collinson G, Combi M, Connerney J, Connour K, Correia J, Crabb K, Crary F, Cravens T, Crismani M, Delory G, Dewey R, DiBraccio G, Dong C, Dong Y, Dunn P, Egan H, Elrod M, England S, Eparvier F, Ergun R, Eriksson A, Esman T, Espley J, Evans S, Fallows K, Fang X, Fillingim M, Flynn C, Fogle A, Fowler C, Fox J, Fujimoto M, Garnier P, Girazian Z, Groeller H, Gruesbeck J, Hamil O, Hanley KG, Hara T, Harada Y, Hermann J, Holmberg M, Holsclaw G, Houston S, Inui S, Jain S, Jolitz R, Kotova A, Kuroda T, Larson D, Lee Y, Lee C, Lefevre F, Lentz C, Lo D, Lugo R, Ma Y-J, Mahaffy P, Marquette ML, Matsumoto Y, Mayyasi M, Mazelle C, McClintock W, McFadden J, Medvedev A, Mendillo M, Meziane K, Milby Z, Mitchell D, Modolo R, Montmessin F, Nagy A, Nakagawa H, Narvaez C, Olsen K, Pawlowski D, Peterson W, Rahmati A, Roeten K, Romanelli N, Ruhunusiri S, Russell C, Sakai S, Schneider N, Seki K, Sharrar R, Shaver S, Siskind DE, Slipski M, Soobiah Y, Steckiewicz M, Stevens MH, Stewart I, Stiepen A, Stone S, Tenishev V, Terada N, Terada K, Thiemann E, Tolson R, Toth G, Trovato J, Vogt M, Weber T, Withers P, Xu S, Yelle R, Yiğit E, Zurek R (2018) Loss of the Martian atmosphere to space: present-day loss rates determined from MAVEN observations and integrated loss through time. *Icarus* 315:146–157
- Krasnopolsky VA (2012) Search for methane and upper limits to ethane and SO₂ on Mars. *Icarus* 217(1):144–152
- Kasting JF (1991) CO₂ condensation and the climate of early Mars. *Icarus* 94:1–13
- Kasprzak WT, Hedin AE, Mayr HG, Niemann HB (1988) Wavelike perturbations observed in the neutral thermosphere of Venus. *J Geophys Res* 93:11237–11245
- Khayat AS, Villanueva GL, Mumma MJ, Tokunaga AT (2015) A search for SO₂, H₂S and SO above Tharsis, and Syrtis volcanic districts on Mars using ground-based high-resolution submillimeter spectroscopy. *Icarus* 253:130–141
- Korablev OI, Montmessin F, Fedorova AA, Ignatiev NI, Shakun AV, Trokhimovskiy AV, Grigoriev AV, Anufreichik KA, Kozlova TO (2015) ACS experiment for atmospheric studies on “ExoMars-2016” orbiter. *Sol Syst Res* 49(7):529–537
- Krasnopolsky VA (2000) On the deuterium abundance on Mars and some related problems. *Icarus* 148:597–602. <https://doi.org/10.1006/icar.2000.6534>
- Krasnopolsky VA, Mumma MJ, Gladstone GR (1998) Detection of atomic deuterium in the upper atmosphere of Mars. *Science* 280:1576
- Krasnopolsky VA, Maillard JP, Owen TC (2004) Detection of methane in the martian atmosphere: evidence for life? *Icarus* 172:537–547
- Leblanc F, Witasse O, Winningham J, Brain D, Liliensten J, Bletly PL, Frahm RA, Halekas JS, Bertaux JL (2006) Origins of the Martian aurora observed by Spectroscopy for Investigation of Characteristics of the Atmosphere of Mars (SPICAM) on board Mars express. *J Geophys Res* 111:A09313. <https://doi.org/10.1029/2006JA0117632>
- Leblanc F, Modolo R, Curry S, Luhmann J, Lillis R, Chaufray JY, Hara T, McFadden J, Halekas J, Eparvier F, Larson D, Connerney J, Jakosky B (2015) Mars heavy ion precipitating flux

- as measured by Mars Atmosphere and Volatile Evolution. *Geophys Res Lett* 42:9135–9141. <https://doi.org/10.1002/2015GL066170>
- Lefèvre F, Forget F (2009) Observed variations of methane on Mars unexplained by known atmospheric chemistry and physics. *Nature* 460:720. <https://doi.org/10.1038/nature08228>
- Lillis RJ, Brain DA, Bougher SW, Leblanc F, Luhmann JG, Jakosky BM, Modolo R, Fox J, Deighan J, Fang X, Wang YC, Lee Y, Dong C, Ma Y, Cravens T, Andersson L, Curry SM, Schneider N, Combi M, Stewart I, Clarke J, Grebowsky J, Mitchell DL, Yelle R, Nagy AF, Baker D, Lin RP (2015) *Space Sci Rev* 195:357–422. <https://doi.org/10.1007/s11214-0165-8>
- Luhmann JG, Johnson RE, Zhang MHG (1992) Evolutionary impact of sputtering of the martian atmosphere by O⁺ pickup ions. *Geophys Res Lett* 19:2151–2154
- Lundin R (2011) Ion acceleration and outflow from Mars and Venus: an overview. *Space Sci Rev* 162:309–334
- Lundin R, Barabash S, Holmström M, Nilsson H, Yamauchi M, Dubinin EM, Fraenz M (2009) *Geophys Res Lett* 36:L17202. <https://doi.org/10.1029/GL039341>
- Ma Y, Nagy AF, Sokolov IV, Hansen KC (2004) Three-dimensional, multispecies, high spatial resolution MHD studies of the solar wind interaction with Mars. *J Geophys Res* 109:A07211. <https://doi.org/10.1029/2003JA010367>
- Maltagliati L, Montmessin F, Fedorova A, Korablev O, Forget F, Bertaux JL (2011) Evidence of water vapor in excess of saturation in the atmosphere of Mars. *Science* 333:1868. <https://doi.org/10.1126/science.1207957>
- Maltagliati L, Montmessin F, Korablev O, Fedorova A, Forget F, Määttänen A, Lefèvre F, Bertaux JL (2013) Annual survey of water vapor vertical distribution and water-aerosol coupling in the martian atmosphere observed by SPICAM/ME_x solar occultations. *Icarus* 223:942–962
- Medvedev AS, Yiğit E (2012) Thermal effects of internal gravity waves in the Martian upper atmosphere. *Geophys Res Lett* 39:L05201. <https://doi.org/10.1029/2012GL50852>
- Medvedev AS, Yiğit E, Hartogh P, Becker E (2011) Influence of gravity waves on the Martian atmosphere: general circulation modeling. *J Geophys Res* 116:E10004. <https://doi.org/10.1029/2011JE003848>
- Medvedev AS, Nakagawa H, Mockel C, Yiğit E, Kuroda T, Hartogh P, Terada K, Terada N, Seki K, Schneider NM, Jain SK, Evans JS, Deighan JI, McClintock WE, Lo D, Jakosky BM (2016) Comparison of the Martian thermospheric density and temperature from IUVS/MAVEN data and general circulation modeling. *Geophys Res Lett* 43:3095–3104
- Montmessin F, Fouchet T, Forget F (2005) Modeling the annual cycle of HDO in the Martian atmosphere. *J Geophys Res* 110:E03006. <https://doi.org/10.1029/2004JE002357>
- Mumma M, Novak RE, DiSanti MA, Bonev BP (2003) *Bulletin of the American Astronomical Society*, 35, AAS/Division for Planetary Sciences Meeting Abstract#35, 937
- Mumma M, Villanueva GL, Novak RE, Hewagama T, Bonev BP, DiSanti MA, Mandell AM, Smith MD (2009) Strong release of methane on Mars in northern summer 2003. *Science* 323:1041. <https://doi.org/10.1126/science.1165243>
- Owen T (1992) In: Kieffer HH (ed) *The composition and early history of the atmosphere of Mars*. University of Arizona Press, Tucson
- Owen T, Maillard JP, Bergh C, Lutz BL (1988) Deuterium on Mars: the abundance of HDO and the value of D/H. *Science* 240(4860):1767. <https://doi.org/10.1126/science.240.4860.1767>
- Parker TJ, Gorsline DS, Saunders RS, Pieri DC, Schneeberger DM (1993) Coastal geomorphology of the martian Northern Plains. *J Geophys Res* 98:11061–11078
- Pollack JB, Kasting JF, Richardson SM, Poliakov K (1987) The case for a wet, warm climate on early Mars. *Icarus* 71(2):203–224
- Ramirez RM, Kopparapu R, Zuger ME, Robinson TD, Freedman R, Kasting JF (2014) *Nat Geosci*. <https://doi.org/10.1038/NGE2000>
- Schneider NM, Deighan JI, Jain SK, Stiepen A, Stewart AIF, Larson D, Mitchell DL, Mazelle C, Lee CO, Lillis RJ, Evans JS, Brain D, Stevens MH, McClintock WE, Chaffin MS, Crismani M, Holsclaw GM, Lefèvre F, Lo DY, Clarke JT, Montmessin F, Jakosky BM (2015) Discovery of diffuse aurora on Mars. *Science* 350:aad0313–aad0311

- Shizgal BD, Arkos GG (1996) Nonthermal escape of the atmospheres of Venus, Earth, and Mars. *Rev Geophys* 34:483–505
- Som SM, Catling DC, Harnmeijer JP, Polivka PM, Buick R (2012) Air density 2.7 billion years ago limited to less than twice modern levels by fossil raindrop imprints. *Nature* 484:359. <https://doi.org/10.1038/nature10890>
- Terada N, Leblanc F, Nakagawa H, Medvedev AS, Yiğit E, Kuroda T, Hara T, England SL, Fujiwara H, Terada K, Seki K, Mahaffy PR, Elrod M, Benna M, Grebowsky J, Jakosky BM (2017) Global distribution and parameter dependences of gravity wave activity in the Martian upper atmosphere derived from MAVEN/NGISM observations. *J Geophys Res* 122:2374–2397. <https://doi.org/10.1002/2016JA023476>
- Vandaele AC, Neefs E, Drummond R, Thomas IR, Daerden F, Lopez-Moreno JL, Rodriguez J, Patel MR, Bellucci G, Allen M, Altieri F, Bolsée D, Clancy T, Delanoye S, Depiesse C, Cloutis E, Fedorova A, Formisano V, Funke B, Fussen D, Geminale A, Gérard JC, Giuranna M, Ignatiev N, Kaminski J, Karatekin O, Lefèvre F, López-Puertas M, López-Valverde M, Mahieux A, McConnell J, Mumma M, Neary L, Renotte E, Ristic B, Robert S, Smith M, Trokhimovsky S, Vander Auwera J, Villanueva G, Whiteway J, Wilquet V, Wolff M (2015) The NOMAD Team, science objectives and performances of NOMAD, a spectrometer suite for the ExoMars TGO mission. *Planet Space Sci* 119:233–249
- Villanueva GL, Mumma MJ, Novak RE, Radeva YL, Käufel HU, Smette A, Tokunaga A, Khayat A, Encrenaz T, Hartogh P (2013) A sensitive search for organics (CH₄, CH₃OH, H₂CO, C₂H₆, C₂H₂, C₂H₄), hydroperoxyl (HO₂), nitrogen compounds (N₂O, NH₃, HCN) and chlorine species (HCl, CH₃Cl) on Mars using ground-based high-resolution infrared spectroscopy. *Icarus* 223:11–27
- Villanueva GL, Mumma MJ, Novak RE, Käufel HU, Hartogh P, Encrenaz T, Tokunaga A, Khayat A, Smith MD (2015) Strong water isotopic anomalies in the martian atmosphere: probing current and ancient reservoirs. *Science*. <https://doi.org/10.1126/science.aaa3630>
- Webster CR, Mahaffy PR, Flesch GJ, Niles PB, Jones JH, Leshin LA, Atreya SK, Stern JC, Christensen LE, Owen T, Franz H, Pepin RO, Steele A (2013) The MSL Science Team, isotope ratios of H, C, and O in CO₂ and H₂O of the Martian atmosphere. *Science* 341:260. <https://doi.org/10.1126/science.1237961>
- Webster CR, Mahaffy PR, Atreya SK, Flesch GJ, Mischna MA, Meslin PY, Farley KA, Conrad PG, Christensen LE, Pavlov AA, Martin-Torres J, Zorzano MP, McConnochie TH, Owen T, Eigenbrode JL, Glavin DP, Steele A, Malespin CA, Archer PD Jr, Sutter B, Coll P, Freissiet C, McKay CP, Moores JE, Schwenzer SP, Bridges JC, Navarro-Gonzalez R, Gellert R, Lemmon MT, the MSL Science Team (2014) Mars methane detection and variability at Gale crater. *Science*. <https://doi.org/10.1126/science/1261713>
- Wordsworth R, Forget F, Millour E, Head JW, Madeleine J-B, Charnay B (2013) Global modelling of the early martian climate under a denser CO₂ atmosphere: water cycle and ice evolution. *Icarus* 222(1):1–19
- Wordsworth RD, Kerber L, Pierrehumbert RT, Forget F, Head JW (2015) Comparison of “warm and wet” and “cold and icy” scenarios for early Mars in a 3-D climate model. *J Geophys Res* 120:1201–1219. <https://doi.org/10.1002/2015JE004787>
- Wordsworth R, Kalugina Y, Lokshtanov S, Vigasin A, Ehlmann B, Head J, Sanders C, Wang H (2017) Transient reducing greenhouse warming on early Mars. *Geophys Res Lett* 44:665–671. <https://doi.org/10.1002/2016GL071766>
- Zahnle K (2015) Play it again. *SAM Sci* 347:370

Chapter 23

The Search for Life on Mars



Yoshitaka Yoshimura

Abstract Though Mars is a cold and dry planet now, Mars would have harbored a large amount of liquid water on the surface early in its history. Mars could have been similar to the early Earth from which life arose 4 billion years ago, and life may have also emerged on Mars during this period. Although the Viking mission in 1976, which explored life on Mars, did not find evidence for life, many findings associated with the possibility of life have been discovered since the Viking mission: past and present aqueous environments, organic compounds, methane, reduced compounds suitable for microorganism energy sources, and so on. These findings suggest that life might exist on Mars. Habitable environments may be deep subsurface, but it may also be on or near the surface where physical and chemical conditions on which even terrestrial microorganisms to survive are found. Life detection instruments have been developed since the Viking mission. Traces or existence of Martian life might be found by future exploration.

Keywords Habitability · Organic compounds · Living microorganisms · Chemolithoautotrophs · Life detection instruments

23.1 Introduction

The possibility of life on Mars has enamored many people for many years. In the 1970s, the Viking landers conducted search-for-life experiments and failed to detect evidence of life on the planet. After the Viking mission, both the National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA) focused on investigating ancient habitability in their Mars exploration programs. Evidence of past liquid water activities has been reported: large outflow channels found by the Mars Global Surveyor (Malin and Carr 1999; Malin and Edgett 2000), H₂O ice under tens of centimeters of soil found by the Mars Odyssey Neutron Spectrometer (Feldman et al. 2002), hydrated sulfate and phyllosilicates found by the Mars

Y. Yoshimura (✉)
Tamagawa University, Tokyo, Japan
e-mail: ystk@agr.tamagawa.ac.jp

Express spacecraft (Gendrin et al. 2005), sedimentary rocks found by the Mars Exploration Rover Opportunity (Squyres and Knoll 2005), and an H₂O ice table found by the Phoenix lander (Smith et al. 2009). These findings suggest that ancient Mars sustained large quantities of liquid water and contained suitable conditions for life (see Chap. 21; Lasue et al. 2013). Assuming that life emerges in suitable conditions, it can be hypothesized that, since life arose within a few hundred million years after the formation of Earth's surface (McKay and Davis 1991), life could have also arisen on early Mars.

Even on Mars today, locations indicating the possible presence of liquid water have been found. Furthermore, organic compounds and energy sources used by terrestrial microorganisms have been discovered on the surface. This chapter reviews the possibility of life on Mars and the potential for exploration.

23.2 The Viking Mission and Recent Reexaminations

NASA conducted life exploration experiments with the Viking mission in 1976; however, the existence of life could not be verified from the surface soil samples (Klein 1977, 1978, 1979, 1992; Margulis et al. 1979). The Viking landers carried out three biological experiments on the Mars surface: the pyrolytic release (PR) experiment for detecting carbon assimilation, the labeled release (LR) experiment for detecting the decomposition of organic compounds, and the gas exchange (GEX) experiment for detecting changes in gas composition caused by metabolic reactions (Fig. 23.1).

In the PR experiment, carbon dioxide and carbon monoxide labeled with radioactive carbon-14 (¹⁴C) were added to the soil sample with water and irradiated with light. In the presence of organisms, ¹⁴C is incorporated, and organic compounds are produced. These compounds were pyrolyzed, and the released ¹⁴C was measured with a radiation detector. Although small amounts of CO₂/CO were incorporated, this incorporation was considered non-biological, because similar levels of incorporation were seen after heating the samples at 90 °C for 2 h.

In the GEX experiment, a nutrient medium containing organic substances, such as amino acids and vitamins, and a mixed gas composed of CO₂ and Kr (in He) were added to the sample chamber. The changes in the gas composition were analyzed by gas chromatography after several days. If organisms were present, the gas composition would be changed by the release of carbon dioxide. Although the CO₂ evolution was observed, it was thought to have come from the oxidation of organics in the nutrient medium by indigenous oxidants like Fe₂O₃ (Oyama and Berdahl 1977).

The LR experiment was conducted by adding seven liquid nutrients (formate, glycolate, glycine, D-alanine, L-alanine, D-lactate, and L-lactate) labeled with ¹⁴C, to the samples. The release of radioactive carbon (such as ¹⁴CO₂) was expected, if organisms metabolized the nutrients. The results showed positive responses that were consistent with biological activities: radioactive gas was evolved, and the gas evolution was reduced or not observed when samples were heated at 46 °C or

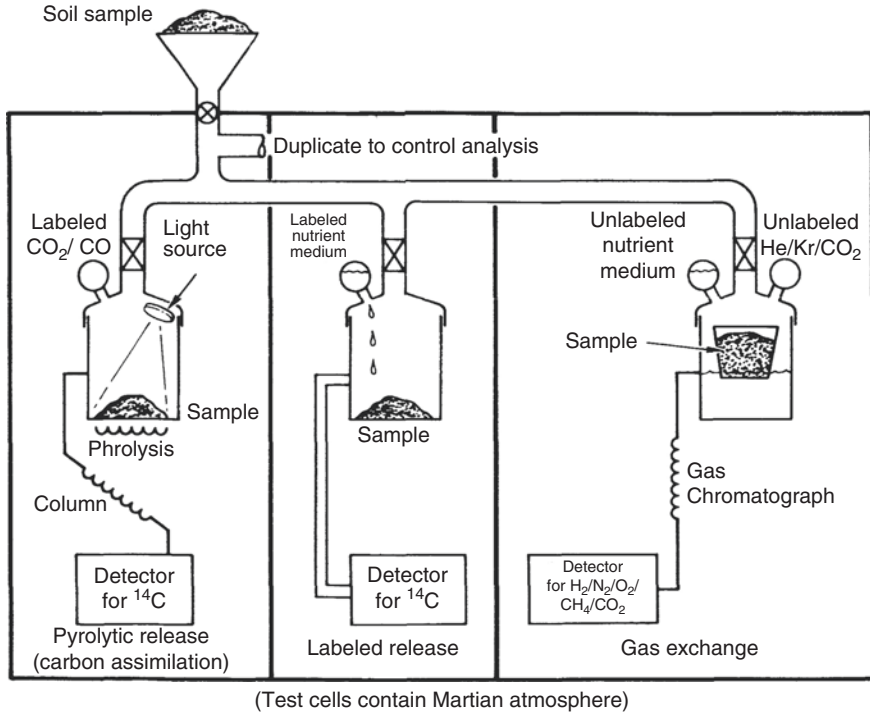


Fig. 23.1 Schematic diagram of the three biological experiments conducted by the Viking landers (Adapted from Viking 1 Early Results, NASA SP-408, 1976)

160 °C, respectively (Levin and Straat 1977). However, the results were still interpreted to be non-biological responses. This interpretation was based on the following results: (1) organic compounds were not detected at levels above the detection limit of the thermal volatilization–gas chromatography–mass spectrometry (TV-GCMS) instrument, and, although chlorinated organics (chloromethane and dichloromethane) were detected, they were interpreted as terrestrial contamination (Biemann et al. 1977); (2) the results could be explained by the presence of oxidants in the regolith (Klein 1978; Margulis et al. 1979). Therefore, the most acceptable conclusion of the Viking experiments was that no organisms were present within the detection limits of these experiments (Klein 1977, 1998, 1999).

However, after the Viking mission, instrumental limitations were reported. The Viking TV-GCMS was not specifically designed to identify the presence of living cells, and the pyrolysis products of cells would not have been detected if living cells were present in quantities less than 10^7 cells per gram (Glavin et al. 2001). Nonvolatile salts of organic acids and low levels of organic compounds would not have been easily detected by the TV-GCMS (Benner et al. 2000; Navarro-Gonzalez et al. 2006). Thus, the existence of organic compounds on Mars could not be accurately determined by the Viking instrument.

23.3 Habitability

In the years since the Viking mission, both NASA and ESA have looked for evidence of ancient habitability, such as traces of past water activities. An ancient possible habitable environment was discovered by the Curiosity Rover at Yellowknife Bay in Gale Crater (Grotzinger et al. 2014). The site was determined to be an ancient lake, with neutral pH and low salinity. Reduced iron and sulfur, as possible microbial energy sources, as well as biogenic elements (C, H, O, S, N, P), have also been detected. On modern-day Mars, life (most likely microorganisms) might exist at least locally since organic compounds, possible liquid water, and energy sources have been found. Figure 23.2 provides a schematic drawing of the possible habitability of present-day Mars. To learn more about both past and present habitability on Mars, see a review by Cockell (2014).

23.3.1 Organic Compounds

If life existed on Mars, organic compounds would be present. The detection and interpretation of organic compounds on Mars are complicated by the existence of perchlorates. While the Viking TV-GCMS did not identify any organic compounds, the Phoenix lander conducted a chemical analysis and detected the presence of 0.4–0.6 wt% perchlorate anion (ClO_4^-) from the surface soil (Hecht et al. 2009). Perchlorate is rare on the surface of Earth; it occurs naturally in hyperarid environments, such as

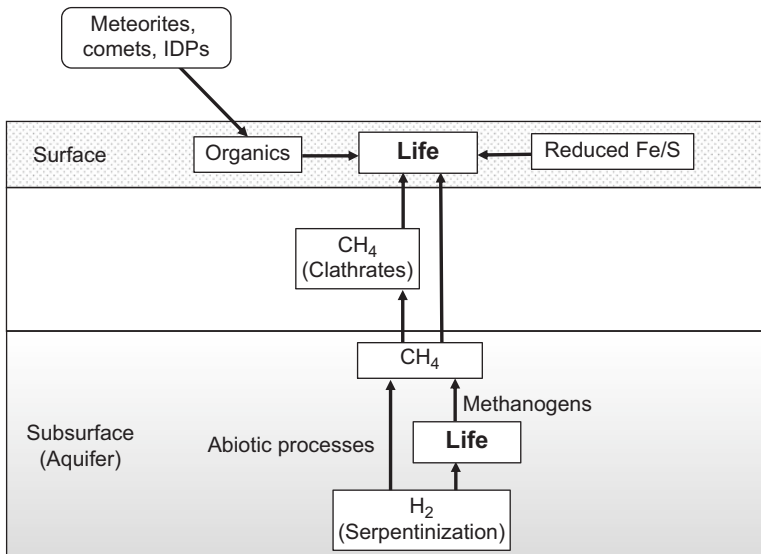


Fig. 23.2 Schematic diagram of possible habitability on modern-day Mars

the Atacama Desert in Chile (Catling et al. 2010) and the Antarctic Dry Valleys (Kounaves et al. 2010). Though perchlorate salts are stable at low temperature, they become strong oxidants when heated, decomposing organic compounds.

Recently, the Curiosity Rover detected chlorinated compounds (chlorobenzene and C2 to C4 dichloroalkanes) (Ming et al. 2014; Freissinet et al. 2015) and thiophenic, aromatic, and aliphatic compounds (Eigenbrode et al. 2018) in the mudstones in Gale Crater. The chlorinated compounds were interpreted to be the reaction products of pyrolysis between oxychlorine compounds, such as perchlorate, and indigenous organic compounds (Freissinet et al. 2015). However, identification of the original organic compounds is difficult due to the complex reactions during pyrolysis. Thus, it is uncertain whether these compounds are derived from Martian sources (igneous, hydrothermal, atmospheric, or biological) or exogenous sources (meteorites, comets, or interplanetary dust particles (IDPs)) (Freissinet et al. 2015). Due to the discovery of perchlorates, it has been noted that the chlorinated compound found by the Viking TV-GCMS might also be the reaction product of indigenous organic compounds and perchlorates during pyrolysis of the soil samples at 500 °C (Steininger et al. 2012; Lasne et al. 2016). Although the discovery of organic compounds does not indicate evidence of life, it also does not rule out the possibility of life. The identification and characterization of organic compounds will be important for future Mars missions.

23.3.2 *Liquid Water*

Liquid water is a fundamental requirement for life. As described in Chap. 21, ancient Mars had a large amount of liquid water on the surface, but liquid water is unstable on the present Martian surface because of low temperatures and pressures. Water on the surface exists mainly in the form of ice, which organisms cannot use. Ground ice has been identified near the surface of the planet by orbiting neutron detectors on the gamma-ray spectrometer carried by Mars Odyssey (Feldman et al. 2002), and the Phoenix lander has shown a shallow H₂O ice table at depths of 5–18 cm in the northern arctic region (Smith et al. 2009).

The possibility of liquid water has also been reported on the Martian surface. For example, recurring slope lineae (RSL), narrow dark streaks on steep slopes that appear during warm seasons in equatorial regions, could be a result of liquid water flow (McEwen et al. 2014) as hydrated salts of magnesium perchlorate, magnesium chlorate, and sodium perchlorate were observed at some of the flows sites (Ojha et al. 2015), although, as another interpretation, it could be dry granular flows of sand and dust (Dundas et al. 2017). The salt solutions can lower the freezing point and the evaporation rate of water. For example, highly concentrated perchlorate solutions remain at liquid state below about –70 °C (Möhlmann and Thomsen 2011). In addition to RSL, the Curiosity Rover demonstrated that transit liquid brine is formed at night in the shallow subsurface at Gale Crater, based on the meteorological analysis. Recently the Mars Advanced Radar for Subsurface and Ionosphere Sounding (MARSIS) instrument on the Mars Express spacecraft has detected the

evidence of stable liquid water, possibly perchlorate brines, about 1.5 km below the surface at the southern polar ice cap (Orosei et al. 2018). Liquid brines could be abundant on Mars, since perchlorates are widespread on the surface (Martín-Torres et al. 2015). Although it is uncertain whether the Martian brines have sufficient water activity to support life (Martín-Torres et al. 2015; Edwards and Piqueux 2016), microbial growth on Earth is known to occur in highly concentrated salt solutions (Grant 2004), and some halophilic microorganisms can use perchlorate as an electron acceptor for respiration in anaerobic conditions (Oren et al. 2014). It can be inferred, therefore, that microorganisms may survive in the briny environments on Mars.

23.3.3 Energy Sources

Life requires energy for its growth, reproduction, and survival. Terrestrial organisms that obtain chemical energy from the oxidation of reduced compounds (energy sources) are named *chemotrophs*, while organisms that use light as an energy source are named *phototrophs*. Chemotrophs are further classified into chemoheterotrophs, which use organic compounds as energy sources, and chemolithoautotrophs, which use inorganic compounds as energy sources. Electron transport from reduced compounds (electron donors) to oxidative compounds (electron acceptors) generates energy for the organisms, which can be calculated as Gibbs free energy. There are many combinations of electron donors and electron acceptors in terrestrial microorganisms, some of which may be used by Martian microorganisms.

Table 23.1 provides examples of potential energy sources for chemotrophs on Mars, which are known to exist or are strongly inferred to exist on the planet. The metabolism of these sources, including H_2 , CH_4 , S^0 , S^{2-} , Fe^{2+} , CO , and organic compounds, has been confirmed in terrestrial microorganisms (Cockell 2014; Rummel et al. 2014; Westall et al. 2015; Cockell et al. 2016). Potential electron acceptors in

Table 23.1 Examples of potential energy sources for chemotrophic life on Mars (Adapted from Rummel et al. 2014; Cockell 2014; Westall et al. 2015)

Energy sources	Electron acceptor	Metabolism
H_2	CO_2	Methanogenesis, acetogenesis
H_2	Fe^{3+} , SO_4^{2-} , S^0 , ClO_4^-	Hydrogen oxidation
CH_4	NO_3^- , Fe^{3+} , MnO_2 , SO_4^{2-}	Methane oxidation
S^0 , S^{2-}	NO_3^- , Fe^{3+} , MnO_2	Sulfur oxidation
Fe^{2+}	NO_3^- , MnO_2	Iron oxidation
CO	CO_2 ,	Methanogenesis, acetogenesis
CO	NO_3^- , H_2O , SO_4^{2-} , ClO_4^-	Carbon monoxide oxidation
Organics	NO_3^- , Fe^{3+} , SO_4^{2-} , ClO_4^-	Anaerobic organics oxidation
Organics	Organics	Fermentation

These substances are known to exist or are strongly inferred to exist on Mars. The redox couples have been confirmed in terrestrial microorganisms

anaerobic conditions are CO_2 , Fe^{3+} , SO_4^{2-} , S^0 , NO_3^- , MnO_2 , ClO_4^- , H_2O , and organic compounds. Oxygen that is produced by photolysis/radiolysis of water might be used as an electron acceptor, although the amount of O_2 is much lower in the Martian atmosphere than on Earth (Westall et al. 2015). Low concentration of molecular hydrogen (H_2) was detected in the upper atmosphere by the space-based telescopes from Earth, which was likely produced by photolysis of water vapor (Krasnopolsky and Feldman 2001). Although H_2 has not been directly measured on Mars surface yet, it is inferred from the presence of olivine and serpentine (Oze and Sharma 2005; Schulte et al. 2006). Fe-bearing minerals and elemental sulfur have been identified on Mars surface (Morris et al. 2007), as well as reduced sulfur such as sulfides (Ming et al. 2014). Nitrate has not been directly detected yet, but detection of NO by the Curiosity Rover suggests the possible presence of nitrate (Stern et al. 2015). Indigenous organics may also be energy sources for chemoheterotrophs, although their structures and accessibility are unknown.

Among these energy sources, methane is a molecule with special interest, because it can be an energy source for methane-oxidizing microorganisms (*methanotrophs*). Methane generation is associated with microbial activities on Earth, where around 80% of natural emissions of methane originate from living microorganisms (Etiope et al. 2011).

Methane in the Martian atmosphere has been reported using a variety of methods (see Chap. 22): Earth-based telescopic observations (Krasnopolsky et al. 1997; Krasnopolsky et al. 2004; Mumma et al. 2009), the Planetary Fourier Spectrometer on board the ESA Mars Express (Formisano et al. 2004; Geminale et al. 2011) ranging from several to tens of parts per billion by volume (ppbv), and the tunable laser spectrometer in the Sample Analysis at Mars on the Curiosity Rover at ~ 7.2 ppbv (Webster et al. 2015). Spatial and seasonal variations of methane (Mumma et al. 2009; Webster et al. 2018), combined with its relatively short lifetime in the Martian atmosphere of about 300 years (Krasnopolsky et al. 2004) and potentially less than 200 days (Lefevre and Forget 2009), indicate that methane has been released into the atmosphere locally and periodically. Though the origins of this methane are uncertain, several generation processes have been proposed, including biotic (microbial) and abiotic processes.

Biotic methane is produced by microorganisms called methanogens, which are anaerobic ones belong to the domain Archaea. Most methanogens use H_2 as an energy source and CO_2 for a carbon source, and some methanogens use CO, acetate, methanol, etc. as energy sources. An important H_2 origin could be subsurface serpentinization (Atreya et al. 2007). Serpentinization is a reaction of olivine- and pyroxene-rich rocks with liquid water, liberating H_2 in the process (Schulte et al. 2006). Carbon dioxide may be derived from the atmosphere, magma degassing, or the thermal decomposition of carbonates on Mars (Oehler and Etiope 2017).

Hydrogen is used as an energy source not only for methanogens but also for a wide variety of chemolithoautotrophic microorganisms on Earth (Table 23.1). Since higher temperatures and pressures would sustain liquid water stably at depths below a few kilometers, a microbial community may exist in the Martian subsurface (Chapelle et al. 2002; Clifford et al. 2010; Michalski et al. 2013).

Possible abiotic methane production mechanisms include geological productions like hydrogeochemical Fischer-Tropsch-type (FTT) reactions after serpentinization (Oze and Sharma 2005); thermogenesis of organics delivered to Mars by meteorites, IDPs, or possible biotic organics (Etiope et al. 2011); geothermal reactions at high temperatures (Oehler and Etiope 2017); ultraviolet degradation of meteoritic organics (Keppler et al. 2012); production by the impact of comets (Krasnopolsky 2006); and volcanic degassing (Atreya et al. 2007). Among them, FTT reactions, which produce methane from the reaction of H_2 with CO_2 , could be important, since they are major abiotic producers of methane on Earth that occur over a wide range of temperatures (<100 to ~500 °C) (Etiope and Sherwood Lollar 2013).

Biotic or abiotic methane could be released into the atmosphere directly and/or via *clathrates* (methane-hydrates) or gas-absorbing regolith. Thus, the presence of methane today does not require the presence of living methanogens; it may have been produced by past methanogens and preserved (Max and Clifford 2000; Atreya et al. 2007).

23.3.4 *Physical and Chemical Conditions*

Terrestrial microorganisms inhabit a wide range of environmental conditions (see Chap. 20). Although present Martian environments are hostile to life, some microorganisms may survive near the surface (Yamagishi et al. 2010). Microorganisms isolated from a Siberian permafrost sample, for example, were capable of growth under simulated Mars conditions: low temperature (0 °C), low pressure (7 hPa), and an anoxic CO_2 -dominated atmosphere (Nicholson et al. 2013).

Radiation would be a serious limiting factor for microbial survivability, since organic compounds are likely to be destroyed by ionizing radiation and UV radiation (Benner et al. 2000; Kminek and Bada 2006). The total dose of ionizing radiation on the Martian surface was measured as 76 mGy/year by the Curiosity Rover (Hassler et al. 2013). However, a radiation-tolerant microbe, *Deinococcus radiodurans*, can survive 5 kGy without loss of viability (Cox and Battista 2005; Dartnell et al. 2007); thus, ionizing radiation would not seriously damage these microorganisms. Though UV radiation is harmful, it would be shielded by thin layers (less than a millimeter) of dust or regolith (Mancinelli and Klovstad 2000). A depth of several centimeters from the surface, therefore, could provide sufficient covering for microorganisms to survive.

23.4 Life Detection Instruments and Possible Explorations

There are many biosignatures for the targets of life explorations, including organic compounds, metabolic activities, cell-like morphology, and stable isotope patterns. Organic compounds are important targets, and the GCMS is an effective instrument

for the detection of organic compounds. However, as mentioned earlier in this chapter, analyses by the TV-GCMS on Mars were affected by perchlorates, which react with indigenous organics during pyrolysis. In forthcoming missions, including the Mars 2020 mission and the ExoMars 2020 mission, instruments designed to detect organic compounds without pyrolysis have been selected.

The Mars 2020 rover will detect organic compounds with the Scanning Habitable Environments with Raman and Luminescence for Organics and Chemicals (SHERLOC), which will detect and characterize minerals and organic compounds, such as aromatic hydrocarbons, with a resonance Raman spectrometer and a fluorescence spectrometer that utilizes a deep-UV laser (<250 nm) (Abbey et al. 2017). It has a context imager with a spatial resolution of 30 μm to visualize surface textures, morphology, and visible features correlated with the spectral signatures (Beegle et al. 2015). The rover will also select and cache the highest value samples for a future sample-return mission, which will take the samples to laboratories on Earth for advanced analysis.

The ExoMars 2020 rover will be equipped with a drill to collect materials from outcrops at depths down to 2 m. The organic compounds will be detected by the Mars Organic Molecule Analyzer (MOMA), which includes two different types of analysis methods, laser desorption mass spectrometry (LD-MS) and TV-GCMS, with or without derivatization agents (Vago et al. 2017). The LD-MS method is not affected by perchlorates, and the derivatization process will be useful for detecting refractory molecules like carboxylic and amino acids. It will be also equipped with a Raman laser spectrometer that will identify minerals and organic compounds (Vago et al. 2017).

The candidates for landing sites in both the Mars 2020 and ExoMars 2020 missions are places with evidence of past water activities (Ono et al. 2016; Kereszturi et al. 2016). Although RSL where possible liquid water/brine exists could indicate attractive sites at which to search for extant life, RSL are not considered indicators for high-priority sites for either project. Both missions have mainly focused on investigating *ancient* habitability; additionally, explorations of RSL would include Committee on Space Research (COSPAR) Planetary Protection constraints to protect Mars from contamination from terrestrial organisms (Rummel and Conley 2017). The current COSPAR Planetary Protection Policy defines Mars special regions as locations in which Earth life could propagate, where the temperature is at or above $-28\text{ }^{\circ}\text{C}$ and water activity is at least 0.5 (Kminek and Rummel 2015; Rummel and Conley 2017). Although no confirmed special regions have been shown on Mars, RSL indicate possible candidates (Rummel et al. 2014; Rummel and Conley 2017). Exhaustive discussion will be required for explorations in RSL areas.

Other attractive sites where to search for present life include methane seepage sites. Even though methanogens and other microorganisms may exist in the deep subsurface, it is difficult to explore those areas. Methanotrophs, however, may be found near the surface (Yamagishi et al. 2010). Some methanotrophs on Earth utilize MnO_2 , $\text{Fe}(\text{OH})_3$, and SO_4^{2-} as electron acceptors (Beal et al. 2009), all of which have been found on the Martian surface. Potential methane seepage sites have been indicated on Mars, such as mud volcano-like mounds, ancient springs, and rims of large impact craters (Oehler and Etiope 2017). When future work by the Trace Gas

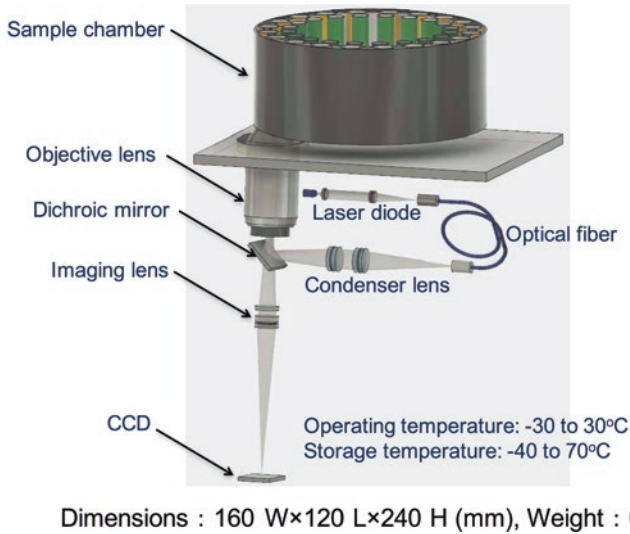


Fig. 23.3 Conceptual design of LDM (Life Detection Microscope) (Adapted from Yamagishi et al. 2018 (© 2018 by the Japan Society for Aeronautical and Space Sciences and ISTS))

Orbiter or surface rovers determines the locations of methane seepage sites, those sites could become candidates for future exploration.

A microscopic instrument would be a powerful tool for searching for extant microorganisms, but it has not been used in space missions yet (Nadeau et al. 2008; Yamagishi et al. 2010). Microscopes directly image life forms and identify their shapes, sizes, and other morphological structures. The Life Detection Microscope (Fig. 23.3) proposed by Yamagishi et al. (2018) detects organic compounds at a spatial resolution of 1 μm , differentiating among organic compounds surrounded by membranes or with enzyme activity by staining the samples with fluorescent pigments. This technique is especially useful for the detection of living microorganisms.

The search for living microorganisms is important not only for scientific interest but for planetary protection. Before future human missions begin, surveys investigating the presence of living microorganisms should be conducted to mitigate the risk of human contact with Martian microorganisms, which may be harmful to human health. Microscopic instruments would be effective tools for this purpose.

23.5 Conclusions

Although the Viking mission failed to detect Martian life, recent findings, such as the presence of organic compounds, energy sources, and possible liquid water, have suggested the possibility of life near the planet's surface. Life detection instruments have been developed since the Viking mission that use Raman spectrometry and

LD-MS without pyrolysis to avoid the problem caused by the reaction between indigenous organic compounds and perchlorates. Microscopic instruments that are particularly superior for detecting living microorganisms have also been proposed. These in situ instruments and a future sample-return mission might reveal the existence of life on Mars.

References

- Abbey WJ, Bhartia R, Beegle LW, DeFlores L, Paez V, Sijapati K, Sijapati S, Williford K, Tuite M, Hug W (2017) Deep UV Raman spectroscopy for planetary exploration: the search for in situ organics. *Icarus* 290:201–214
- Atreya SK, Mahaffy PR, Wong A-S (2007) Methane and related trace species on Mars: origin, loss, implications for life, and habitability. *Planet Space Sci* 55(3):358–369
- Beal EJ, House CH, Orphan VJ (2009) Manganese- and iron-dependent marine methane oxidation. *Science* 325(5937):184–187
- Beegle L, Bhartia R, White M, DeFlores L, Abbey W, Wu Y-H, Cameron B, Moore J, Fries M, Burton A (2015) SHERLOC: scanning habitable environments with Raman & luminescence for organics & chemicals. In *Aerospace Conference, 2015 IEEE*, pp 1–11
- Benner SA, Devine KG, Matveeva LN, Powell DH (2000) The missing organic molecules on Mars. *Proc Natl Acad Sci U S A* 97(6):2425–2430
- Biemann K, Oro J, Toulmin PIII, Orgel LE, Nier AO, Anderson DM, Simmonds PG, Flory D, Diaz AV, Rushneck DR, Biller JE, Lafleur AL (1977) The search for organic substances and inorganic volatile compounds in the surface of Mars. *J Geophys Res* 82(28):4641–4658. <https://doi.org/10.1029/JS082i028p04641>
- Catling D, Claire M, Zahnle K, Quinn R, Clark B, Hecht M, Kounaves S (2010) Atmospheric origins of perchlorate on Mars and in the Atacama. *J Geophys Res: Planets* 115(E1):E00E11
- Chapelle FH, O’neill K, Bradley PM, Methé BA, Ciuffo SA, Knobel LL, Lovley DR (2002) A hydrogen-based subsurface microbial community dominated by methanogens. *Nature* 415(6869):312–315
- Clifford SM, Lasue J, Heggy E, Boisson J, McGovern P, Max MD (2010) Depth of the Martian cryosphere: revised estimates and implications for the existence and detection of subpermafrost groundwater. *J Geophys Res: Planets* 115(E7):E07001
- Cockell CS (2014) Trajectories of Martian habitability. *Astrobiology* 14(2):182–203
- Cockell C, Bush T, Bryce C, Direito S, Fox-Powell M, Harrison J, Lammer H, Landenmark H, Martin-Torres J, Nicholson N (2016) Habitability: a review. *Astrobiology* 16(1):89–117
- Cox MM, Battista JR (2005) *Deinococcus radiodurans* – the consummate survivor. *Nat Rev Microbiol* 3(11):882–892
- Dartnell LR, Desorgher L, Ward J, Coates A (2007) Modelling the surface and subsurface Martian radiation environment: implications for astrobiology. *Geophys Res Lett* 34(2):L02207
- Dundas CM, McEwen AS, Chojnacki M, Milazzo MP, Byrne S, McElwaine JN, Urso A (2017) Granular flows at recurring slope lineae on Mars indicate a limited role for liquid water. *Nat Geosci* 10(12):903–907
- Edwards CS, Piqueux S (2016) The water content of recurring slope lineae on Mars. *Geophys Res Lett* 43(17):8912–8919
- Eigenbrode JL, Summons RE, Steele A, Freissinet C, Millan M, Navarro-González R, Sutter B, McAdam AC, Franz HB, Glavin DP, Archer PD, Mahaffy PR, Conrad PG, Hurowitz JA, Grotzinger JP, Gupta S, Ming DW, Sumner DY, Szopa C, Malespin C, Buch A, Coll P (2018) Organic matter preserved in 3-billion-year-old mudstones at Gale crater. *Mar Sci* 360(6393):1096–1101
- Etioppe G, Sherwood Lollar B (2013) Abiotic methane on Earth. *Rev Geophys* 51(2):276–299

- Etiopio G, Oehler D, Allen C (2011) Methane emissions from Earth's degassing: implications for Mars. *Planet Space Sci* 59(2):182–195
- Feldman WC, Boynton WV, Tokar RL, Prettyman TH, Gasnault O, Squyres SW, Elphic RC, Lawrence DJ, Lawson SL, Maurice S, McKinney GW, Moore KR, Reedy RC (2002) Global distribution of neutrons from Mars: results from Mars odyssey. *Science* 297(5578):75–78
- Formisano V, Atreya S, Encrenaz T, Ignatiev N, Giuranna M (2004) Detection of methane in the atmosphere of Mars. *Science* 306(5702):1758–1761. <https://doi.org/10.1126/science.1101732>
- Freissinet C, Glavin DP, Mahaffy PR, Miller KE, Eigenbrode JL, Summons RE, Brunner AE, Buch A, Szopa C, Archer PD, Franz HB, Atreya SK, Brinckerhoff WB, Cabane M, Coll P, Conrad PG, Marais DJD, Dworkin JP, Fairén AG, François P, Grotzinger JP, Kashyap S, Ilt K, Leshin LA, Malespin CA, Martin MG, Martin-Torres FJ, McAdam AC, Ming DW, Navarro-González R, Pavlov AA, Prats BD, Squyres SW, Steele A, Stern JC, Sumner DY, Sutter B, Zorzano MP (2015) Organic molecules in the Sheepbed Mudstone, Gale Crater, Mars. *J Geophys Res: Planets* 120(3):495–514. <https://doi.org/10.1002/2014JE004737>
- Geminale A, Formisano V, Sindoni G (2011) Mapping methane in Martian atmosphere with PFS-MEX data. *Planet Space Sci* 59(2):137–148
- Gendrin A, Mangold N, Bibring JP, Langevin Y, Gondet B, Poulet F, Bonello G, Quantin C, Mustard J, Arvidson R, LeMouélic S (2005) Sulfates in Martian layered terrains: the OMEGA/Mars express view. *Science* 307(5715):1587–1591
- Glavin DP, Schubert M, Botta O, Kminek G, Bada JL (2001) Detecting pyrolysis products from bacteria on Mars. *Earth Planet Sci Lett* 185(1–2):1–5
- Grant W (2004) Life at low water activity. *Philos Trans R Soc Lond B: Biol Sci* 359(1448):1249–1267
- Grotzinger JP, Sumner DY, Kah L, Stack K, Gupta S, Edgar L, Rubin D, Lewis K, Schieber J, Mangold N (2014) A habitable fluvio-lacustrine environment at Yellowknife Bay, Gale Crater. *Mar Sci* 343(6169):1242777
- Hassler DM, Zeitlin C, Wimmer-Schweingruber RF, Ehresmann B, Rafkin S, Eigenbrode JL, Brinza DE, Weigle G, Böttcher S, Böhm E (2013) Mars' surface radiation environment measured with the Mars Science Laboratory's Curiosity rover. *Science* 343:1244797
- Hecht MH, Kounaves SP, Quinn RC, West SJ, Young SM, Ming DW, Catling DC, Clark BC, Boynton WV, Hoffman J, Deflores LP, Gospodinova K, Kapit J, Smith PH (2009) Detection of perchlorate and the soluble chemistry of Martian soil at the Phoenix lander site. *Science* 325(5936):64–67. <https://doi.org/10.1126/science.1172466>
- Keppeler F, Vigano I, McLeod A, Ott U, Früchtl M, Röckmann T (2012) Ultraviolet-radiation-induced methane emissions from meteorites and the Martian atmosphere. *Nature* 486(7401):93
- Kereszturi A, Bradák B, Chatzitheodoridis E, Ujvari G (2016) Indicators and methods to understand past environments from ExoMars rover drills. *Orig Life Evol Biospheres* 46(4):435–454
- Klein HP (1977) The Viking biological investigation: general aspects. *J Geophys Res* 82(28):4677–4680
- Klein HP (1978) The Viking biological experiments on Mars. *Icarus* 34(3):666–674
- Klein HP (1979) The Viking mission and the search for life on Mars. *Rev Geophys* 17(7):1655–1662
- Klein HP (1992) The Viking biology experiments: epilogue and prologue. *Orig Life Evol Biospheres* 21(4):255–261
- Klein HP (1998) The search for life on Mars: what we learned from Viking. *J Geophys Res* 103(E12):28463–28466. <https://doi.org/10.1029/98je01722>
- Klein HP (1999) Did Viking discover life on Mars? *Orig Life Evol Biospheres* 29(6):625–631
- Kminek G, Bada JL (2006) The effect of ionizing radiation on the preservation of amino acids on Mars. *Earth Planet Sci Lett* 245(1–2):1–5
- Kminek G, Rummel J (2015) COSPAR's planetary protection policy. *Space Res Today* 193:7–18
- Kounaves SP, Stroble ST, Anderson RM, Moore Q, Catling DC, Douglas S, McKay CP, Ming DW, Smith PH, Tamppari LK (2010) Discovery of natural perchlorate in the Antarctic Dry Valleys and its global implications. *Environ Sci Technol* 44(7):2360–2364

- Krasnopolsky VA (2006) Some problems related to the origin of methane on Mars. *Icarus* 180(2):359–367
- Krasnopolsky VA, Feldman PD (2001) Detection of molecular hydrogen in the atmosphere of Mars. *Science* 294(5548):1914–1917
- Krasnopolsky V, Bjoraker G, Mumma M, Jennings D (1997) High-resolution spectroscopy of Mars at 3.7 and 8 μm : a sensitive search for H_2O_2 , H_2CO , HCl , and CH_4 , and detection of HDO . *J Geophys Res: Planets* 102(E3):6525–6534
- Krasnopolsky VA, Maillard JP, Owen TC (2004) Detection of methane in the Martian atmosphere: evidence for life? *Icarus* 172(2):537–547
- Lasne J, Noblet A, Szopa C, Navarro-González R, Cabane M, Poch O, Stalport F, François P, Atreya SK, Coll P (2016) Oxidants at the surface of Mars: a review in light of recent exploration results. *Astrobiology* 16(12):977–996
- Lasue J, Mangold N, Hauber E, Clifford S, Feldman W, Gasnault O, Grima C, Maurice S, Mousis O (2013) Quantitative assessments of the Martian hydrosphere. *Space Sci Rev* 174(1–4):155–212
- Lefevre F, Forget F (2009) Observed variations of methane on Mars unexplained by known atmospheric chemistry and physics. *Nature* 460(7256):720
- Levin GV, Straat PA (1977) Recent results from the Viking labeled release experiment on Mars. *J Geophys Res* 82(28):4663–4667. <https://doi.org/10.1029/JS082i028p04663>
- Malin MC, Carr MH (1999) Groundwater formation of Martian valleys. *Nature* 397(6720):589–591. <https://doi.org/10.1038/17551>
- Malin MC, Edgett KS (2000) Sedimentary rocks of early Mars. *Science* 290(5498):1927–1937
- Mancinelli RL, Klovstad M (2000) Martian soil and UV radiation: microbial viability assessment on spacecraft surfaces. *Planet Space Sci* 48(11):1093–1097
- Margulis L, Mazur P, Barghoorn ES, Halvorson HO, Jukes TH, Kaplan IR (1979) The Viking Mission: implications for life on Mars. *J Mol Evol* 14(1):223–232
- Martín-Torres FJ, Zorzano M-P, Valentín-Serrano P, Harri A-M, Genzer M, Kempainen O, Rivera-Valentín EG, Jun I, Wray J, Madsen MB (2015) Transient liquid water and water activity at Gale crater on Mars. *Nat Geosci* 8(5):357–361
- Max MD, Clifford SM (2000) The state, potential distribution, and biological implications of methane in the Martian crust. *J Geophys Res: Planets* 105(E2):4165–4171
- McEwen AS, Dundas CM, Mattson SS, Toigo AD, Ojha L, Wray JJ, Chojnacki M, Byrne S, Murchie SL, Thomas N (2014) Recurring slope lineae in equatorial regions of Mars. *Nat Geosci* 7(1):53–58
- McKay CP, Davis WL (1991) Duration of liquid water habitats on early Mars. *Icarus* 90(2):214–221
- Michalski JR, Cuadros J, Niles PB, Parnell J, Rogers AD, Wright SP (2013) Groundwater activity on Mars and implications for a deep biosphere. *Nat Geosci* 6(2):133
- Ming D, Archer P, Glavin D, Eigenbrode J, Franz H, Sutter B, Brunner A, Stern J, Freissinet C, McAdam A (2014) Volatile and organic compositions of sedimentary rocks in Yellowknife Bay, Gale Crater. *Mar Sci* 343(6169):1245267
- Möhlmann D, Thomsen K (2011) Properties of cryobrine on Mars. *Icarus* 212(1):123–130
- Morris R, Ming D, Yen A, Arvidson R, Gruener J, Humm D, Klingelhöfer G, Murchie S, Schröder C, Seelos IV F, Squyres S., Wisema S., Wolff M., the MER and CRISM Science Teams (2007) Possible evidence for iron sulfates, iron sulfides, and elemental sulfur at Gusev Crater, Mars, from MER, CRISM, and analog data. In *Seventh International Conference on Mars*
- Mumma MJ, Villanueva GL, Novak RE, Hewagama T, Bonev BP, Disanti MA, Mandell AM, Smith MD (2009) Strong release of methane on Mars in northern summer 2003. *Science* 323(5917):1041–1045. <https://doi.org/10.1126/science.1165243>
- Nadeau JL, Perreault NN, Niederberger TD, Whyte LG, Sun HJ, Leon R (2008) Fluorescence microscopy as a tool for in situ life detection. *Astrobiology* 8(4):859–874. <https://doi.org/10.1089/ast.2007.0043>

- Navarro-Gonzalez R, Navarro KF, de la Rosa J, Iniguez E, Molina P, Miranda LD, Morales P, Cienfuegos E, Coll P, Raulin F, Amils R, McKay CP (2006) The limitations on organic detection in Mars-like soils by thermal volatilization-gas chromatography-MS and their implications for the Viking results. *Proc Natl Acad Sci U S A* 103(44):16089–16094. <https://doi.org/10.1073/pnas.0604210103>
- Nicholson WL, Krivushin K, Gilichinsky D, Schuerger AC (2013) Growth of *Carnobacterium* spp. from permafrost under low pressure, temperature, and anoxic atmosphere has implications for Earth microbes on Mars. *Proc Natl Acad Sci* 110(2):666–671. <https://doi.org/10.1073/pnas.1209793110>
- Oehler DZ, Etiope G (2017) Methane seepage on Mars: where to look and why. *Astrobiology* 17(12):1233–1264. <https://doi.org/10.1089/ast.2017.1657>
- Ojha L, Wilhelm MB, Murchie SL, McEwen AS, Wray JJ, Hanley J, Masse M, Chojnacki M (2015) Spectral evidence for hydrated salts in recurring slope lineae on Mars. *Nat Geosci* 8:829–832. <https://doi.org/10.1038/ngeo2546>
- Ono M, Throck B, Almeida E, Ansar A, Otero R, Huertas A, Heverly M (2016) Data-driven surface traversability analysis for Mars 2020 landing site selection. In: *Aerospace Conference, IEEE*, pp 1–12
- Oren A, Bardavid RE, Mana L (2014) Perchlorate and halophilic prokaryotes: implications for possible halophilic life on Mars. *Extremophiles* 18(1):75–80
- Orsei R, Lauro SE, Pettinelli E, Cicchetti A, Coradini M, Cosciotti B, Di Paolo F, Flamini E, Mattei E, Pajola M, Soldovieri F, Cartacci M, Cassenti F, Frigeri A, Giuppi S, Martufi R, Masdea A, Mitri G, Nenna C, Noschese R, Restano M, Seu R (2018) Radar evidence of subglacial liquid water on Mars. *Science* 361:eaar7268
- Oyama VI, Berdahl BJ (1977) The Viking gas exchange experiment results from Chryse and Utopia surface samples. *J Geophys Res* 82(28):4669–4676. <https://doi.org/10.1029/JS082i028p04669>
- Oze C, Sharma M (2005) Have olivine, will gas: serpentinization and the abiogenic production of methane on Mars. *Geophys Res Lett* 32(10):L10203
- Rummel J, Conley C (2017) Four fallacies and an oversight: searching for Martian life. *Astrobiology* 17(10):1–4
- Rummel JD, Beaty DW, Jones MA, Bakermans C, Barlow NG, Boston PJ, Chevrier VF, Clark BC, de Vera J-PP, Gough RV (2014) A new analysis of Mars “special regions”: findings of the second MEPAG Special Regions Science Analysis Group (SR-SAG2). *Astrobiology* 14(11):887–968
- Schulte M, Blake D, Hoehler T, McCollom T (2006) Serpentinization and its implications for life on the early Earth and Mars. *Astrobiology* 6(2):364–376
- Smith PH, Tappari LK, Arvidson RE, Bass D, Blaney D, Boynton WV, Carswell A, Catling DC, Clark BC, Duck T, DeJong E, Fisher D, Goetz W, Gunnlaugsson HP, Hecht MH, Hipkin V, Hoffman J, Hviid SF, Keller HU, Kounaves SP, Lange CF, Lemmon MT, Madsen MB, Markiewicz WJ, Marshall J, McKay CP, Mellon MT, Ming DW, Morris RV, Pike WT, Renno N, Staufer U, Stoker C, Taylor P, Whiteway JA, Zent AP (2009) H₂O at the phoenix landing site. *Science* 325(5936):58–61. <https://doi.org/10.1126/science.1172339>
- Squyres SW, Knoll AH (2005) Sedimentary rocks at Meridiani Planum: origin, diagenesis, and implications for life on Mars. *Earth Planet Sci Lett* 240(1):1–10
- Steininger H, Goemann F, Goetz W (2012) Influence of magnesium perchlorate on the pyrolysis of organic compounds in Mars analogue soils. *Planet Space Sci* 71(1):9–17
- Stern JC, Sutter B, Freissinet C, Navarro-González R, McKay CP, Archer PD, Buch A, Brunner AE, Coll P, Eigenbrode JL (2015) Evidence for indigenous nitrogen in sedimentary and aeolian deposits from the Curiosity rover investigations at Gale crater, Mars. *Proc Natl Acad Sci* 112(14):4245–4250
- Vago JL, Westall F, Coates AJ, Jaumann R, Korablev O, Ciarletti V, Mitrofanov I, Josset J-L, De Sanctis MC, Bibring J-P (2017) Habitability on early Mars and the search for biosignatures with the ExoMars Rover. *Astrobiology* 17(6–7):471–510

- Webster CR, Mahaffy PR, Atreya SK, Flesch GJ, Mischna MA, Meslin P-Y, Farley KA, Conrad PG, Christensen LE, Pavlov AA (2015) Mars methane detection and variability at Gale crater. *Science* 347(6220):415–417
- Webster CR, Mahaffy PR, Atreya SK, Moores JE, Flesch GJ, Malespin C, McKay CP, Martinez G, Smith CL, Martin-Torres J, Gomez-Elvira J, Zorzano M-P, Wong MH, Trainer MG, Steele A, Archer D, Sutter B, Coll PJ, Freissinet C, Meslin P-Y, Gough RV, House CH, Pavlov A, Eigenbrode JL, Glavin DP, Pearson JC, Keymeulen D, Christensen LE, Schwenger SP, Navarro-Gonzalez R, Pla-García J, Rafkin SCR, Vicente-Retortillo Á, Kahanpää H, Viudez-Moreiras D, Smith MD, Harri A-M, Genzer M, Hassler DM, Lemmon M, Crisp J, Sander SP, Zurek RW, Vasavada AR (2018) Background levels of methane in Mars' atmosphere show strong seasonal variations. *Science* 360(6393):1093–1096
- Westall F, Foucher F, Bost N, Bertrand M, Loizeau D, Vago JL, Kminek G, Gaboyer F, Campbell KA, Bréhéret J-G (2015) Biosignatures on Mars: what, where, and how? Implications for the search for Martian life. *Astrobiology* 15(11):998–1029
- Yamagishi A, Yokobori S, Yoshimura Y, Yamashita M, Hashimoto H, Kubota T, Yano H, Haruyama J, Tabata M, Kobayashi K, Honda H, Utsumi Y, Saiki T, Itoh T, Miyakawa A, Hamase K, Naganuma T, Mita H, Tonokura K, Sasaki S, Miyamoto H (2010) Japan Astrobiology Mars Project (JAMP): search for microbes on the Mars surface with special interest in methane-oxidizing bacteria. *Biol Sci Space* 24(2):67–82
- Yamagishi A, Satoh T, Miyakawa A, Yoshimura Y, Sasaki S, Kobayashi K, Kebukawa Y, Yabuta H, Mita H, Imai E, Naganuma T, Fujita K, Usui T (2018) LDM (life detection microscope): in situ imaging of living cells on surface of Mars. *Transaction of the Japan Society for Aeronautical and Space Sciences. Aerosp Technol Jpn* 16(ISTS31):299–305

Chapter 24

Active Surface and Interior of Europa as a Potential Deep Habitat



Jun Kimura

Abstract Jupiter's moon Europa may have an internal ocean of liquid water, along with the chemical compounds and energy source that life requires. Europa is covered by the solid icy shell, similar to other solid bodies in the outer solar system. The solid icy shell fractures and deforms creating cracks, ridges, and bands in relatively a recent period. Galileo spacecraft data indicates a warm interior, which means a convecting icy shell above a liquid water ocean. In addition, Hubble Space Telescope recently found a signature of active water plumes from the southern hemisphere. Here the current knowledge on the characteristic of Europa, geology, composition, interior, and surrounding environment, in the relation to the possible presence of life will be summarized. Future spacecraft exploration plans for Europa and their science objectives are also introduced. With the understanding of Europa's potential for life, we can consider another style of habitable world hidden by the icy surface, "deep habitat," which is different from Earth's one, and can address the fundamental question: Are we alone in the universe?

Keywords Satellite of Jupiter · Ice · Habitability · Tectonics · Interior

24.1 Introduction and History

Europa, the smallest of the Galilean satellites with a radius of 1565 km (Fig. 24.1), orbits around Jupiter with a period of 3.55 days at an average orbital radius of 6.71×10^5 km (9.4 Jovian radius) and is gravitationally locked to Jupiter such that the same hemisphere of the moon always faces the planet. The gravitationally locked Europa has leading and trailing hemispheres regarding the orbiting direction. Europa is in a tug-of-war with Io and Ganymede, and Europa's orbital period is twice Io's period and half of Ganymede's one. In other words, every time Ganymede goes around Jupiter once, Europa makes two orbits, and Io makes four orbits. This is a well-known example of a 1:2:4 orbital resonance (also called Laplace

J. Kimura (✉)
Osaka University, Toyonaka City, Osaka, Japan
e-mail: junkim@ess.sci.osaka-u.ac.jp

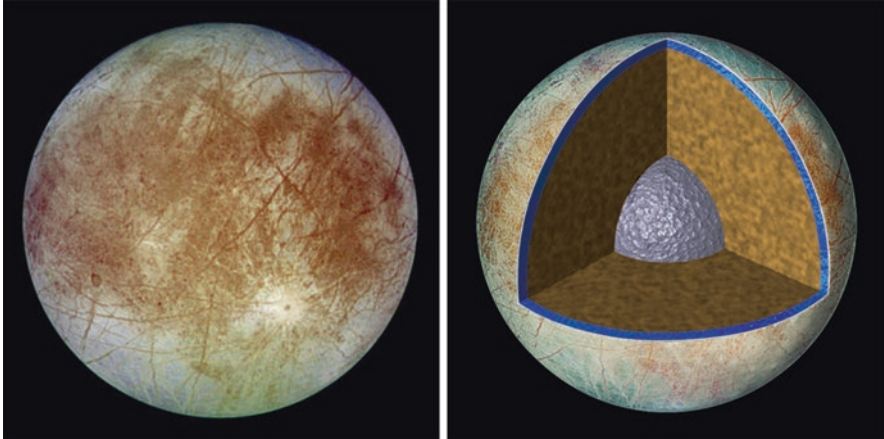


Fig. 24.1 (Left) Natural-color appearance of the trailing hemisphere of Europa taken by Galileo SSI (Solid-State Imager). The diameter of Europa is 1560 km. (Right) Cutaway view of the possible internal structure of Europa (NASA/JPL-Caltech)

Table 24.1 Basic Europa parameters (Weiss 2004)

Quantity	Value
Orbital semi-major axis	671,100 km
Orbital semi-major axis	9.38 Jovian radii
Orbital period	3.551 Earth days
Eccentricity	0.0094
Mean radius	1560.8 ± 0.5 km
Mass	$4.8017 \pm 0.000014 \times 10^{22}$ kg
Bulk density	3014 ± 5 kg/m ³

resonance) in orbiting bodies and thus has a forced nonzero eccentricity (Table 24.1). Tidal deformation and dissipative heating arising from the eccentric orbit can be the source of geological activity in those satellites.

After Galileo Galilei's discovery of Europa (January 1610) with a 1-inch magnifier, the Pioneer 10 took a first spacecraft image of Europa (December 1973). Although it was not a close flyby to the satellites and that the image was of very low resolution, careful tracking of the spacecraft refined the estimation of the mean density for Europa of 2.99 g/cm³. The Voyager "grand tour" followed upon Pioneers 10 and 11. Upon encountering Jupiter, two Voyager spacecrafts (March and July 1979) unveiled Europa's young surface with enigmatic surface features and determined the diameter of Europa and other Galilean moons, which stand until now. Almost 12 years after Voyager's launch in 1977, Galileo spacecraft launched on October 1989, carried by Space Shuttle Atlantis. Galileo arrived at Jupiter on December 1995 after Venus and Earth's gravity assist flybys and became the first spacecraft to orbit Jupiter. However, the deployment of the main 4.8-m high-gain

antenna failed to open fully; less than 1% of the originally planned data could be transmitted to Earth. After epic changes of software and data compression schemes, Galileo encountered Europa 12 times in total and acquired data using 16 instruments, bringing us many findings specifically of unique geology, induced magnetic field, atmosphere and exosphere, surface compositions, and constraints for interior and astrobiological perspectives.

Europa's surface albedo is high, and its global spectrum is compatible with that of clear water ice. Its bulk density of 3.01 g/cm^3 indicates that there must be at least a few percent of water, which is larger than the total amount of terrestrial surface water. Measurements of Europa's gravity field inferred that its interior is differentiated at least into an outer water shell, including both a solid ice crust and a liquid ocean existing under the crust, with 200 km or less thickness and an inner rock-iron core. Magnetic measurements strongly showed that Europa has an induced magnetic field through interaction with Jupiter's field and strongly suggested the presence of a subsurface conductive layer, a possible global ocean of salty liquid water. Furthermore, erupting jets of water vapor from Europa's southern hemisphere have been suggested from images taken by the Hubble Space Telescope. If confirmed, it would open the opportunity of a flyby through the plume and obtain a sample to analyze in situ, as it was done for the Saturnian moon Enceladus.

24.2 Shape, Gravity, Interior

Europa and all other synchronous rotating bodies are distorted into a triaxial ellipsoid in response to the satellite's rotation and the tidal potential due to the host planet. Accurate measurements of the shape can provide constraints for its internal structure. Although direct shape measurements for Europa have not yet been performed, a recent estimate is consistent with a hydrostatic state (Nimmo et al. 2007). With the assumption that Europa is in hydrostatic equilibrium, it is possible to determine values of the gravitational coefficients and to infer its axial moment of inertia. Hydrostaticity has not yet been verified, and it needs to be confirmed quantitatively from independent measurements of Europa's shape using laser altimetry and/or imaging data analysis in future mission.

Today, the interior characteristics of Europa are inferred from gravity field measurements by the Galileo spacecraft. On the assumption that Europa has the hydrostatic equilibrium figure that the shape of Europa's equipotential surface of the gravity results from a physical distortion caused by the tidal forces and by its spin, the figure and gravitational field can be described by a degree-2 spherical harmonic function as a good approximation. Europa's axial moment of inertia C which is inferred by the gravitational field tells us about the distribution of mass in the interior and normalized value by MR^2 (M is the mass of Europa and R is its radius) $C/MR^2 = 0.346 \pm 0.005$. The value for a homogeneous density sphere substantially smaller than 0.4 indicates that the density increases toward the center of Europa. Based on the constraint of the mean density and the moment of inertia, the interior

of Europa has been modeled as a two- or three-layer structure. The two-layer model of Europa composed of an ice outer shell and a silicate/metal-mixed inner core requiring a density of the latter layer higher than 3800 kg/m^3 . Such a higher density of the interior indicates high metal enrichment relative to Io and is unlikely for a smaller body forming further out of Io in the proto-Jovian nebula. In addition, radiogenic heating in the silicates would raise the temperature high enough for the differentiation of metallic component from silicates. Therefore, Europa must have a three-layer structure with an Fe or FeS core at its center, a rocky mantle surrounding the metallic core, and a water shell overlying the rocky mantle (Fig. 24.1, right). The radius of the metallic core is between 40% and 50% of Europa's radius depending on the density of the core, e.g., denser and smaller Fe core, or lighter and larger FeS core. The thickness of the outer water shell is between 80 and 170 km. The gravity data cannot conclude the physical states (i.e., solid or liquid) of the water layer and the metallic core because of the small density differences between these states.

In addition, measurements of magnetic environment around Europa have inferred the presence of a global liquid salty ocean underneath the solid ice shell (Kivelson et al. 2000). Measurement of the Jovian magnetic field when the Galileo spacecraft performed a flyby of Europa has confirmed the signatures that the Jovian magnetic field was disturbed before and after the closest approach to Europa. This disturbance can be explained by the existence of a magnetic dipole tilted about 90° with respect to the rotation axis inside Europa of which direction is changed with the variation of the Jovian magnetosphere. Because the axis of the Jovian magnetosphere is tilted about 10° from its rotation axis, Europa passes through the south and northern hemisphere of the Jovian magnetosphere upon orbiting around Jupiter (Zimmer et al. 2000). Thus, the direction of the Jovian magnetic field applied to Europa itself periodically fluctuates, and the electric conductor existing inside Europa produces eddy currents in response to this fluctuation (Khurana et al. 1998). Accordingly, the fluctuation of the magnetic field suggests the presence of salty ocean in Europa.

24.3 Surface State, Composition, Plume, and Atmosphere

Early spectroscopic studies (e.g., Moroz 1965; Pilcher et al. 1972) suggested that the surface of Europa is largely dominated by water ice. The ice at the very surface is either a fluffy porous ice or is fine-powdered grains of ice, which has been interpreted from diurnal temperature measurements, photometric observations, and sputter modeling (Moore et al. 2009). A fluffy snow-like surface could be the result of precipitation from vapor clouds ejected through cracks in the ice shell and of fragmentation and mixing by micrometeoroid impacts.

Grain size and crystal structure of Europa's surface ice could be affected by temperature variations and surface thermal processes. Equatorial ice on the leading hemisphere is generally fine grained, radii of 20–50 μm (Hansen et al. 2004), but the trailing hemisphere ice is generally larger grained (radii greater than 200 μm

(McCord et al. 1999). The top layer of Europa's ice is predominantly amorphous probably due to irradiation and disruption of the original crystalline structure (Hansen et al. 2004), although interplanetary dust particle impacts could anneal amorphous ice grains back into crystalline grains (Porter et al. 2010).

Albedo and color differences in hemispheric and local scale are remarkable properties. The surface of the leading hemispheres is brighter than that of the trailing hemisphere, which is attributed to more impacts by charged particles from the Jovian magnetosphere. SO_2 has been detected especially on the trailing hemisphere where charged particle flux becomes strongest because charged particles nearly corotate with Jupiter, continually overtaking the satellite in the orbital motion (Lane et al. 1981; McEwen 1986), and sulfur ion implantation is considered to come from Io's volcanic eruption. In a local scale, a reddish-brown material coats fractures and other geologically young features (see next section). Spectroscopic observations suggest that such reddish materials may be rich in salts such as magnesium sulfate or carbonate, deposited by evaporating salty water that seep out from inside (McCord et al. 1999). In addition, sodium chloride is another possible explanation for brownish discoloration. Irradiated sodium chloride with total dose corresponding to 10–100 years on Europa's surface provides a straightforward explanation for the Solid-State Imaging continuum across the visible wavelength range (Hand et al. 2015).

Hubble Space Telescope (HST) observations in December 2012 of ultraviolet emission at Europa have been interpreted as the existence of water vapor plumes near the South Pole (Roth et al. 2014). The height of the plume derived from these observations is ~ 200 km. The observations also indicated a mass flux of 7000 kg/s and source kinetic temperature of the gas of 230 K (Fig. 24.2). Additional imaging evidence from the HST was acquired between January and April 2014 (Sparks et al.

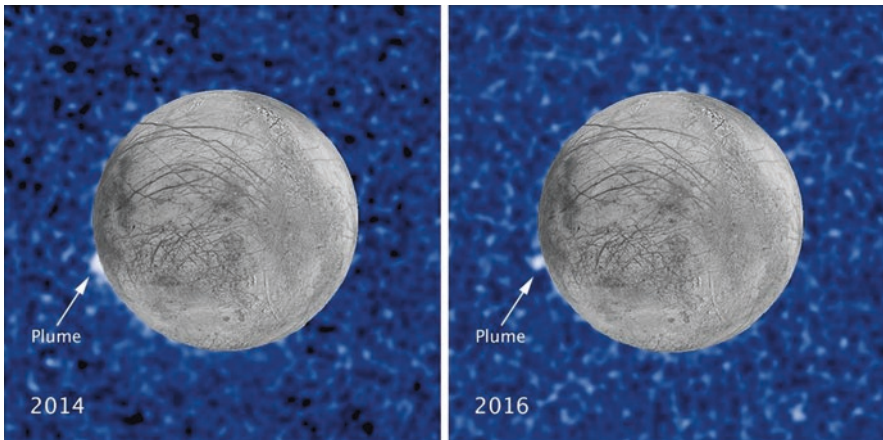


Fig. 24.2 Suspected water plume erupting off the limb of Europa 2 years apart from the same location on Europa, photographed in ultraviolet light by the Hubble Space Telescope Imaging Spectrograph. The image of Europa, superimposed on the Hubble data, is assembled from data from the Galileo and Voyager missions. (NASA/ESA/W. Sparks (STScI)/USGS Astrogeology Science Center)

2016). However, similar HST observations in 1999, June 2008, June 2009, November 2012, and during November 2014 and April 2015 found no plumes. Such observations could imply that the plumes are not persistent but intermittent.

Europa has an extremely tenuous atmosphere with surface pressure of $0.1 \mu\text{Pa}$, 10^{-12} times that of the Earth (Hall et al. 1995). The main component of the atmosphere is molecular oxygen, and it is considered that the surface water ice is dissociated by solar UV and high-energy particles in the Jovian magnetosphere and that oxygen is gravitationally bounded while hydrogen is dissipated.

24.4 Surface Geology

Europa's surface has a notable paucity of large impact craters. The overall trend of Europa's crater size-frequency distributions (SFDs) is largely distinct from the SFDs seen in the inner solar system (e.g., Moon) (Strom et al. 2015). Impacting population of Europa (and other Galilean moons) is comprised mainly of Jupiter-family comets and long-period comets (Levison et al. 2000; Zahnle et al. 2003) and is significantly different from the small asteroids, which come from the main-belt and hit terrestrial bodies in the inner solar system (Strom et al. 2015). Therefore, the production function of impact craters derived from the Moon and Mars are inconsistent with Europa. Europa's average surface age is estimated to be around 20–200 Myr with uncertainties in the comet impact rate and cratering mechanisms (Bierhaus et al. 2009) and thus quite young within the timescale of the solar system. Despite the lack of large craters, each shape and morphology is a window into the interior structure and its activity. The relationship between depth and diameter of impact craters on Europa shows that the depth of craters larger than about 3 km in diameter is shallower than those on the terrestrial moon and that the depth of craters larger than 8 km across decreases with increasing diameter. Such a reverse trend in the depth/diameter relationship is probably due to the process that the excavation cavity induced upon penetrated deep into a very weak and highly mobile layer of the warm part of the ice shell inducing flattening of the larger crater (Schenk and Turtle 2009).

Despite the lack of large craters, numerous and wide variety of geologic features cover the Europa's icy surface, indicating that Europa has experienced significant tectonic disruption of the surface over its history. Surface features on Europa are classified into seven types, including craters explained above, large ringed features, lineae, flexūs, chaos, maculae, and regiones (Fig. 24.3); all classifications are officially named by the International Astronomical Union Working Group on Planetary System Nomenclature. *Large ringed features* show dark and circulate spots with low-rugged and concentric rings with a diameter larger than 100 km and are interpreted as ancient impact craters that have penetrated into deeper warm ice or liquid water. *Lineae* (singular, linea) are defined as any elongate texture with positive or negative topography and are the dominant surface features on Europa. Positive-relief lineae are called as ridges, which are ~200 m to >4 km wide and of 100 s m height and commonly have lengths of >1000 km. Most of the ridges are found in

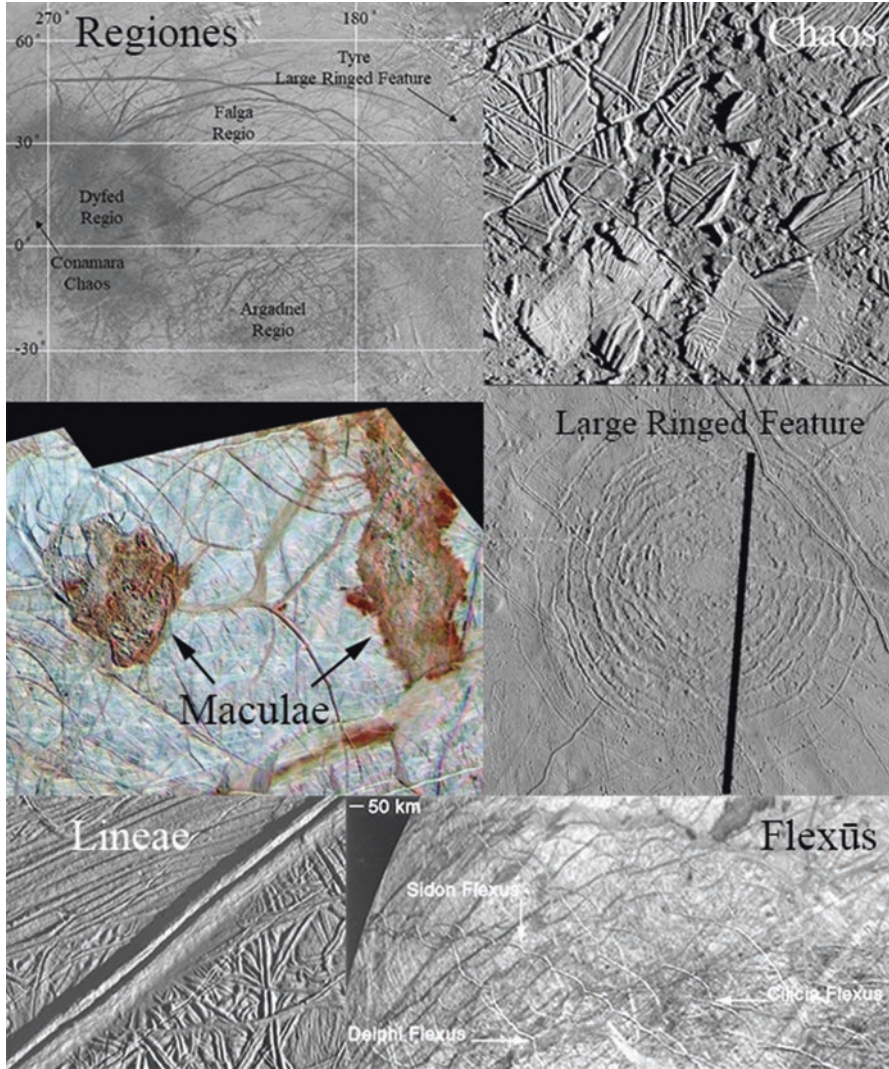


Fig. 24.3 Six types of surface features on Europa (NASA/JPL-Caltech)

double-ridge form, having a central crack or trough flanked by two raised edifices. Regarding the ridge formation mechanism, despite a number of models have been proposed, no single model can explain all the observational facts. In one of the models, the ridges are formed in a similar manner to the terrestrial arctic sea ice. It is suggested that cracks opened by diurnal tidal stresses allow water to seep up from the underlying ocean, filling the crack and freezing into a slurry. Cyclic diurnal stresses would induce compression to the crack, and partially frozen ice is smashed up and squeezed out of the crack to form a pile of ice forming a double ridge on an

initial crack (Greenberg et al. 1998; Tufts et al. 2000). However, there are some arguments against the model: shallow depths of resulting surface fractures (~200 m) because of the small amplitude of the diurnal tidal stress that cannot penetrate the brittle portion (1–3 km thickness) of Europa's ice shell (Greenberg et al. 1998; Lee et al. 2005). In addition, Lee et al. (2005) suggested the width of fractures would be too small, an order of centimeters, to allow the liquid water to seep up from the ocean explaining the model. Another model for the ridge formation suggests that surface fractures undergo shear stress and strike-slip motion (horizontal movement with opposite directions) due to diurnal tides. This motion along the crack induces shear heating sufficient to trigger warm ice to rise up buoyantly to form a ridge (Gaidos and Nimmo 2000; Nimmo and Gaidos 2002). If shear heating is sufficiently large to produce partial melting of ice below the ridge, the melt drains downward and forms the distinct V-shaped central trough of the double ridge. Ridge formation seems to continue throughout Europa's history, while it has been proposed that Europa's ice shell has gradually thicken according to the thermal evolution (Greeley et al. 2004; Kimura et al. 2007). Thus, the dominant ridge formation process perhaps has been changed with time, for example, the diurnal tidal pumping model was predominant when the ice shell is very thin (~ few kms). However, there is no morphological evidence for the varying formation style with time.

Flexūs (singular, flexus) are chains of arcuate ridge segments linked at sharp cusps, which are also called as cycloids. These shapes have been interpreted as curved tensile fractures controlled by a rotating diurnal stress field at an appropriate propagation speed (~3 km/h) slower than Europa's rotation speed (Hoppa et al. 1999).

Chaos is defined as distinctive areas of locally disrupted terrain with polygonal ice blocks of preexisted plains within a matrix of mound-shaped materials, which have been firstly seen in high-resolution Galileo images. Conamara Chaos, a 75 × 100 km disrupted zone, is a typical example and the largest chaos on Europa. Blocks of preexisting terrain have been tilted, rotated, and translated horizontally and are usually and locally higher (~100 m) than adjacent matrix. The matrix is usually domed (~100 to 200 m) above background terrain. Dark hydrated salt of sulfuric acid material is exposed on the matrix, small blocks, and surrounding region (Greeley et al. 2004). Among chaos, small-scale, circular, and elliptical deformed features that are tens to hundreds of meters in positive or negative relief with sizes roughly 10–20 km across in particular are commonly termed as *lenticulae*. Though the formation process of chaos is not fully understood, two models, melt-through and brine mobilization, have been widely accepted and can explain observational characteristics. The melt-through model has been characterized by the melting of a thin conducting ice shell from below due to amplified heat flow through the ocean, from the surface of the rocky mantle to the bottom of the ice shell (Thomson and Delaney 2001; O'Brien et al. 2002). Visual similarity between chaos and terrestrial pack ice leads the model in which ice rafts of preexisting terrain survive within a lumpy matrix and shifted somewhat relative to their original position (Spaun et al. 1998). This model would require a large plume of concentrated heat that is stable for at least hundreds of years to permit transfer of heat from the rocky mantle to the base of the ice shell. If Europa's rocky mantle is highly efficient to dissipate the

kinetic energy from the tidal deformation into heat and generates several terawatts of heat (assuming tidal dissipation in an Io-like mantle), concentrating few percent of the heat could achieve complete melt-through of a 6-km-thick ice shell (O'Brien et al. 2002). However, even if Europa's mantle is such extremely dissipative, hot water from the seafloor is difficult to create a significant number of the most abundant small features with 10 km across or less (*lenticulae*) because heat from the seafloor is difficult to rotationally confine narrow cylinders and delivered heat at the bottom of the ice shell spreads laterally into cones with several tens of kms (Goodman et al. 2004).

The brine mobilization model considers the effect of antifreeze (low melting point) materials and their (partial) melting in the ice shell. On Earth, if sea ice rapidly freezes, salty components can be trapped within their structure. A similar process may occur in Europa's ocean, and exogenic cometary materials can be continuously supplied into the shallow region of the ice shell through the impact. Thus, there are possibly compositional variations on Europa's surface. If an ascending thermal plume in the ice shell approaches the eutectic point of the overlying impure brittle ice, localized lens-shaped partial melting occurs. Drainage of brine from the matrix to other locations will lose the volume compared to the adjacent ice blocks, and thus it makes ice blocks higher than matrix. Furthermore, if the tidal heating concentrates in a thermal plume, enhanced heat generates localized melting even in pure ice. Formation of melt lens causes volume decreasing; thus the surface above the lens subsides, collapses, and calves ice blocks. Brine infiltrates to the resulting crack and forms mound-shaped matrix. Refreezing of the melt lens and freezing of briny matrix raise above the surrounding terrain (Schmidt et al. 2011).

Maculae are dark spots with circular, elliptical, elongate, or irregular shape. Although these were originally interpreted as cryovolcanic flows that flooded a small basin, Galileo high-resolution images showed that they seem to be a chaos-like disrupted terrain, but it is sunken below the surrounding terrain. One of the biggest macula is Thera that is currently sustaining a large liquid lake in the shallow subsurface based on the melt-lens model mentioned above (Schmidt et al. 2011). The formation of *lenticulae* could be caused by the injection of saucer-shaped sills residing several kilometers below the surface of Europa (Manga and Michaut 2017), while cryovolcanism may be responsible for the formation of small domes that do not exhibit the same geological features of their surrounding terrain (Quick et al. 2017)

Regiones are defined as extensive areas where reflectance and color clearly differ from their surroundings based on the low-resolution images taken by the Voyager spacecraft. Subsequent Galileo spacecraft's high-resolution images revealed detailed features characterizing each regio.

Europa's global stress mechanisms, interior structure, and surface geology are inherently linked as discussed above. In addition, analyses of Voyager and Galileo images have found evidence of subduction on Europa's surface, suggesting that new surface area created along expansive wide linear features, commonly called as bands, and physically removed by subduction at the compressional band (Kattenhorn and Prockter 2014). It implies that the surface ice plates analogous to tectonic plates on

Earth may be recycled into the interior of the ice shell and that Europa's ice shell might be broken up into a patchwork of tectonic plates. The main unknowns are the thickness of the icy shell and the ice rheology, which mainly depends on the ice grain size. Constraining the thickness of the icy shell in the future mission and/or modeling is crucial to examine the formation models of geological features on Europa.

24.5 Possible Life on Europa?

Harboring subsurface ocean underneath the ice shell in Europa means that there is an environment where the water-rock interaction, possible hydrothermal circulation, can take place at the seafloor. Because only the low-pressure phase ice (ice Ih) and the liquid phase appear in the low pressure range of H₂O layer in Europa, the solidification of the ocean proceeds only due to the growth of the ice shell above the ocean. In other words, Europa's ocean is always in direct contact with the rocky seafloor, and the decay heat of radioisotopes generated in the rock layer directly heats the ocean. One of the important implications in terms of water-rock interaction is a possible interaction between water and ultramafic rock, called as serpentinization (Vance et al. 2007). In another Jovian moon Io, ultramafic rock that generates hydrogen through an interaction with water has been proposed (Williams et al. 2000; Nna-Mvondo and Martinez-Frias 2007). Analogy can be applied to Europa. Applying Earthlike material properties and cooling rates, the extent in depth of hydrothermal circulation in Europa's seafloor could be ~25 km which is several times larger than in Earth (typically ~6 km), suggesting the possibility of deep biosphere in Europa. Although corresponding heat generation through serpentinization is a few orders of magnitude smaller than the estimated value of tidal heating and radioactive decay heat, generated transition metals such as iron and other substances can be supplied into the ocean. In addition, hydrogen generated due to the serpentinization can be used by living organisms as a reductant (Vance et al. 2007). Therefore, it can be said that serpentinization is an important chemical process in terms of astrobiology.

The surface of the Earth has been oxidative for more than 2 billion years, allowing respiratory organisms to survive. In Europa, however, since the supply of oxidants is perhaps poor, the hydrothermal vents in Europa may not necessarily be an oasis of life. Oxidizing sources can be oxygen O₂, ozone O₃, or hydrogen peroxide H₂O₂, which could be generated from water H₂O due to irradiation with ultraviolet or radiation. Europa's surface is exposed to cosmic radiation (mainly an electron beam) with a flux of 8×10^{13} eV cm⁻² s⁻¹ (Chyba and Phillips 2001), and it forms very tenuous 10⁻¹² atm atmosphere composed of O₂ and O₃ around Europa. In addition, ice on Europa includes hydrogen peroxide at about 0.13% molar ratio to water (Carlson et al. 1999). Assuming that the surface residence time of convective ice is 10 million years, most of the ice components up to 1.3 m in depth become H₂O₂ and O₂ due to the irradiation of space radiation during that period and which could be sent to the ocean through the convection in the ice shell (Chyba and Phillips 2001).

This could be the main oxidants of Europa. In other words, this supply of oxidants will determine the upper limit of biological activity of Europa.

In addition, carbon monoxide CO and carbon dioxide CO₂ are also oxidants. When hydrogen H₂ generated by the serpentinization reacts with CO or CO₂, methane CH₄ is formed. It corresponds to the Sabatier reaction in chemical engineering, and it is also known that this is done by microorganisms (methanogenic bacteria, see Chap. 23). Once methane can be produced, organisms that use it as an energy source and a carbon source can also be reproduced. This is why hydrogen and methane are important for life.

Mineral components could be contained in the subsurface ocean and the melt lens (local lake) in the ice shell, while oxidants such as O₂ and O₃ generated by irradiation could easily go into the lake, which is closer to the surface. In this point, it is conceivable that the possibility of existence of life is higher in the lake, whereas the subsurface ocean is supposed to be poor in oxidants.

24.6 Future Explorations

In 2015, National Aeronautics and Space Administration (NASA) approved the development of a flagship-level mission to explore Europa, named as Europa Clipper, with the specific goal of investigating its habitability. The spacecraft will launch sometime in the next decade and will arrive in the Jupiter system between 3 and 7 years later, depending on the launch vehicle and trajectory. In order to survive the intense Jovian radiation environment, the spacecraft will orbit Jupiter, not Europa, which will dive in and out of the radiation belts. Radiation is expected to be the limiting factor of the mission's lifetime, rather than out of fuel or the pointing disability of instruments in most other spacecraft missions. In the current mission plan, Clipper will make 45 flybys of Europa at different positions and altitudes, ranging from 25 to 2700 km above the moon's surface. The payload consists of five remote sensing instruments that cover the wavelength range from ultraviolet through radar and four in situ instruments that measure fields and particles. Here is the full list of nine science instruments expected on the mission:

- Plasma Instrument for Magnetic Sounding (PIMS)
- Interior Characterization of Europa using MAGnetometry (ICEMAG)
- Mapping Imaging Spectrometer for Europa (MISE)
- Europa Imaging System (EIS)
- Radar for Europa Assessment and Sounding: Ocean to Near-surface (REASON)
- Europa THERmal Emission Imaging System (E-THEMIS)
- MAss SPectrometer for Planetary EXploration/Europa (MASPEX)
- Ultraviolet Spectrograph/Europa (UVS)
- SURface Dust Mass Analyzer (SUDA).

Moreover, gravity science can be achieved via the spacecraft's telecommunication system, and the spacecraft's engineering radiation monitoring system could provide

valuable scientific data. High-priority science will be accomplished through investigations of the moon's interior structure, composition, geology, and current activity.

Currently another spacecraft mission to the Jovian system is under development. JUPITER ICy Moons Explorer (JUICE) is the European Space Agency (ESA)-led mission that will provide the most comprehensive exploration of the Jovian system, specifically addressing two key questions of ESA's Cosmic Vision program: (1) What are the conditions for planet formation and the emergence of life? (2) How does the solar system work (Grasset et al. 2013)? The overarching theme for JUICE is the emergence of habitable worlds around gas giants. The icy Galilean moons of Jupiter – Europa, Ganymede, and Callisto – are believed to contain global subsurface water oceans beneath their icy crusts (e.g., Kivelson et al. 1999, 2000, 2002; Saur et al. 2014). JUICE will uncover the whole picture of Ganymede by the first orbiting in the history around an extraterrestrial moon. In addition, the flybys at Europa and Callisto will deepen our understanding of the current state and evolution of the Jovian satellite system. JUICE is currently planned to be launched in May 2022. Following an interplanetary cruise of 7.6 years, Jupiter orbit insertion will take place in October 2029. The spacecraft will perform a 2.5-year Jupiter-orbiting tour including two flybys of Europa at 400 km altitude and multiple flybys of Ganymede and Callisto with a minimum altitude of 200 km. After these flybys, JUICE will enter into an orbit around Ganymede and stay there for at least 10 months. The payload consists of ten state-of-the-art instruments plus one experiment that uses the spacecraft telecommunication system with ground-based instruments. This payload is capable of addressing all of the mission's science goals, from in situ measurements of Jupiter's atmosphere and plasma environment to remote observations of the surface and interior of the three icy moons, Ganymede, Europa, and Callisto. The following is the list of science instruments expected on the mission:

- Jovis, Amorum ac Natorum Undique Scrutator (JANUS: optical camera system)
- Moons and Jupiter Imaging Spectrometer (MAJIS)
- UV Imaging Spectrograph (UVS)
- Submillimetre Wave Instrument (SWI)
- Ganymede Laser Altimeter (GALA)
- Radar for Icy Moon Exploration (RIME)
- Gravity and Geophysics of Jupiter and the Galilean Moons (3GM)
- Magnetometer for JUICE (J-MAG)
- Particle Environment Package (PEP)
- Radio and Plasma Wave Investigation (RPWI)
- Planetary Radio Interferometer and Doppler Experiment (PRIDE)

24.7 Conclusion

On a giant planet system, only one or two explorations have been performed to date. Spacecraft explorations orbiting icy moons have not yet been achieved, and there are still many big issues to be solved for direct investigation of extraterrestrial life. However, recent observations strongly suggest that the ice shell of Europa is warm in the interior and that the subsurface ocean exists. In addition, active water plumes have been suggested, and if confirmed then it makes possible to identify various organic materials and to evaluate internal energy through direct sampling and analysis of the internal compounds. These facts and implications evoke the existence of the deep habitat driven by satellite's orbital and geothermal energy, which is different in style from Earth and Mars, driven mainly by solar energy. Therefore, we now know the possible place where harbors extraterrestrial life is not limited to the Earthlike habitat at the planetary surface, and the possible habitable zone that has various compounds and energies for life there according to the environment of each bodies. We need to accumulate geological and geochemical knowledge through further explorations and observations, toward discovering the life that may exist in extraterrestrial bodies. In addition, we have to consider carefully what kind of life and evolutionary path to life exist (or existed in the past) in each body and to set up the specific direction and method for upcoming explorations.

References

- Bierhaus EB et al (2009) Europa's crater distributions and surface ages. In: Europa. The University of Arizona Press, Tucson, pp 161–180
- Carlson RW et al (1999) Hydrogen peroxide on the surface of Europa. *Science* 283:2062
- Chyba CF, Phillips CB (2001) Possible ecosystems and the search for life on Europa. *Proc Natl Acad Sci U S A* 98:801–804
- Gaidos EJ, Nimmo F (2000) Tectonics and water on Europa. *Nature* 405:637
- Goodman JC et al (2004) Hydrothermal plume dynamics on Europa: implications for chaos formation. *J Geophys Res* 109. <https://doi.org/10.1029/2003JE002073>
- Grasset O et al (2013) JUPiter ICy moons Explorer (JUICE): an ESA mission to orbit Ganymede and to characterise the Jupiter system. *Planet Space Sci* 78:1–21
- Greeley R et al (2004) Geology on Europa. In: Jupiter: the planet, satellites and magnetosphere. Cambridge University Press, Cambridge, pp 329–362
- Greenberg R et al (1998) Tectonic processes on Europa: tidal stresses, mechanical response, and visible features. *Icarus* 135:64–78
- Hall DT et al (1995) Detection of an oxygen atmosphere on Jupiter's moon Europa. *Nature* 373:677–679
- Hand KP et al (2015) Europa's surface color suggests an ocean rich with sodium chloride. *Geophys Res Lett* 42:3174–3178

- Hansen GB et al (2004) Amorphous and crystalline ice on the Galilean satellites: a balance between thermal and radiolytic processes. *J Geophys Res* 109:E01012. <https://doi.org/10.1029/2003JE002149>
- Hoppa GV et al (1999) Formation of cycloidal features on Europa. *Science* 285:1899–1902
- Kattenhorn SA, Prockter LM (2014) Evidence for subduction in the ice shell of Europa. *Nat Geosci* 7:762–767
- Khurana KK et al (1998) Induced magnetic fields as evidence for subsurface oceans in Europa and Callisto. *Nature* 395:777–780
- Kimura J et al (2007) Tectonic history of Europa: coupling between internal evolution and surface stresses. *Earth Planets Space* 59:113–125
- Kivelson MG et al (1999) Europa and Callisto: induced or intrinsic fields in a periodically varying plasma environment. *J Geophys Res* 104:4609–4625
- Kivelson MG et al (2000) Galileo magnetometer measurements: a stronger case for a subsurface ocean at Europa. *Science* 289:1340–1343
- Kivelson MG et al (2002) The permanent and inductive magnetic moments of Ganymede. *Icarus* 157:507–522
- Lane et al (1981) Evidence for sulfur implantation in Europa's UV absorption band. *Nature* 292:38–39
- Lee S et al (2005) Mechanics of tidally driven fractures in Europa's ice shell. *Icarus* 165:267–379
- Levison et al (2000) NOTE: planetary impact rates from ecliptic comets. *Icarus* 143:415–420
- Manga M, Michaut C (2017) Formation of lenticulae on Europa by saucer-shaped sills. *Icarus* 286:261–269
- McCord TB et al (1999) Hydrated salt minerals on Europa's surface from the Galileo Near Infrared Spectrometer (NIMS) investigation. *J Geophys Res* 104:11,827–11,852
- McEwen A (1986) Exogenic and endogenic albedo and color patterns on Europa. *J Geophys Res* 91:8077–8097
- Moore JM et al (2009) Surface properties, regolith, and landscape degradation. In: *Europa*. The University of Arizona Press, Tucson, pp 329–352
- Moroz VI (1965) Infrared spectroscopy of satellites: the moon and the Galilean satellites of Jupiter. *Astron Zh* 42(1287), translated in *Soviet Astron* 9: 999–1006
- Nimmo F, Gaidos E (2002) Strike-slip motion and double ridge formation on Europa. *J Geophys Res* 107:1–2. <https://doi.org/10.1029/2000JE001476>
- Nimmo F et al (2007) The global shape of Europa: constraints on lateral shell thickness variations. *Icarus* 191:183–192. <https://doi.org/10.1016/j.icarus.2007.04.021>
- Nna-Mvondo D, Martinez-Frias J (2007) Komatiites: from Earth's geological settings to planetary and astrobiological contexts. *Earth Moon Planet* 100:157–179
- O'Brien DP et al (2002) A melt-through model for chaos formation on Europa. *Icarus* 156:152–161
- Pilcher CB et al (1972) The Galilean satellites; identification of water frost. *Science* 178:1087–1089
- Porter SB et al (2010) Micrometeorite impact annealing of ice in the outer solar system. *Icarus* 208:492–498
- Quick LC et al (2017) Cryovolcanic emplacement of domes on Europa. *Icarus* 284:477–488
- Roth L et al (2014) Transient water vapor at Europa's south pole. *Science* 343:171–174
- Saur J et al. (2014) The search for a subsurface ocean in Ganymede with Hubble space telescope observations of its auroral ovals. *J Geophys Res* 120. <https://doi.org/10.1002/2014JA020778>
- Schenk P, Turtle E (2009) Europa's impact craters: probes of the icy shell. In: *Europa*. The University of Arizona Press, Tucson, pp 181–198
- Schmidt BE et al (2011) Active formation of 'chaos terrain' over shallow subsurface water on Europa. *Nature* 479:502–505
- Sparks WB et al (2016) Probing for evidence of plumes on Europa with HST/STIS. *Astrophys J* 829:121
- Spaun NA et al (1998) Conamara Chaos region, Europa: reconstruction of mobile polygonal ice blocks. *Geophys Res Lett* 25:4277

- Strom RG et al (2015) The inner solar system cratering record and the evolution of impactor population. *Astron Astrophys* 15:407–434
- Thomson RE, Delaney JR (2001) Evidence for a weakly stratified European ocean sustained by seafloor heat flux. *J Geophys Res* 106:12355–12365
- Tufts BR et al (2000) Lithospheric dilation on Europa. *Icarus* 146:75–97
- Vance S et al (2007) Hydrothermal systems in small ocean planets. *Astrobiology* 7(6):987–1005
- Weiss JW (2004) Planetary parameters, in *Jupiter: the planet, satellites and magnetosphere*. Cambridge University Press, London, pp 699–709
- Williams DA et al (2000) A komatiite analog to potential ultramafic materials on Io. *J Geophys Res* 105:1671–1684
- Zahnle K et al (2003) Cratering rates in the outer solar system. *Icarus* 163:263–289
- Zimmer C et al (2000) Subsurface oceans on Europa and Callisto: constraints from Galileo magnetometer observations. *Icarus* 147:329–347

Chapter 25

Enceladus: Evidence and Unsolved Questions for an Ice-Covered Habitable World



Yasuhito Sekine, Takazo Shibuya, and Shunichi Kamata

Abstract The icy mid-sized satellite of Saturn—Enceladus—has become the central to astrobiology since the finding of its dramatic ongoing geological activity. The water-rich plumes erupting from the warm fractures on the icy crust near the South Pole of Enceladus originate from its global subsurface ocean that interacts with the rocky core. In situ measurements of the plume by the Cassini spacecraft showed that the ocean contains dissolved gas species, such as CO_2 , NH_3 , CH_4 , and H_2 , which can provide disequilibrium redox energy to support methanogenic life. The salt composition of the plume indicates an alkaline pH of the ocean (pH ~9 to 11). The plume also contains significant amounts of organic matter, including high-molecular-weight organic compounds, although its origin remains unclear. Ongoing hydrothermal activity at temperatures greater than 90°C is highly likely to exist on the seafloor or within the rocky core, which could play a role in sustaining the chemical disequilibrium within the ocean. These observations suggest that Enceladus is a planetary body thus far that currently meets the fundamental requirements for habitability and life—liquid water, organic matter, and bioavailable energy—beyond Earth.

Keywords Icy satellite · Geochemistry · Habitability

Y. Sekine (✉)

Earth-Life Institute, Tokyo Institute of Technology, Meguro, Japan

e-mail: sekine@elsi.jp

T. Shibuya

Department of Subsurface Geobiological Analysis and Research, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka, Japan

S. Kamata

Creative Research Institution, Hokkaido University, Sapporo, Japan

25.1 Introduction

Enceladus is one of the seven regular satellites of Saturn. Its diameter is about 500 km, and its mass ($\sim 1 \times 10^{20}$ kg) is $\sim 0.1\%$ of the largest moon of Saturn, Titan. Since the discovery of the ongoing geological activity on Enceladus by NASA's Cassini mission (e.g., Porco et al. 2006), this moon has become one of the highest-priority targets for astrobiological exploration. The most remarkable geological activity on Enceladus is the water-rich plumes erupting from the warm fractures (termed, tiger stripes) in the South Pole region (e.g., Porco et al. 2006). Multiple lines of Cassini's observations indicate that Enceladus' plumes originate from the subsurface ocean that interacts with the rocky core of the satellite (Fig. 25.1) (e.g., Waite Jr et al. 2009; Postberg et al. 2009; Hsu et al. 2015). The Cassini spacecraft, until its end of mission in 2017, made 22 times of close flybys around Enceladus. During these flybys, the chemical composition of the plumes was investigated using Cassini's onboard payload instruments, e.g., Ion Neutral Mass Spectrometer (INMS), Cosmic Dust Analyzer (CDA), and Ultraviolet Imaging Spectrometer (UVIS) (e.g., Waite Jr et al. 2009, 2017; Postberg et al. 2009, 2011; Hansen et al. 2011). Owing to the results obtained by these instruments together with theoretical modeling and laboratory experiments, Enceladus is known to possess liquid water, organic matter, and bioavailable energy that can support life. In this chapter, we review the geophysical and geochemical processes occurring on Enceladus that could be vital for the habitability on it. We also discuss the unsolved questions that lie ahead for understanding habitability on this small icy moon.

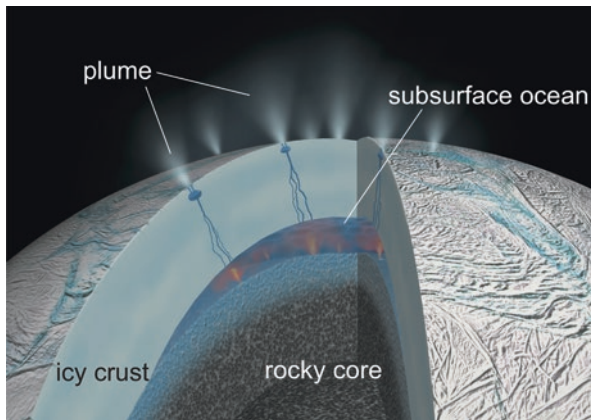


Fig. 25.1 An artist concept of the subsurface ocean of Enceladus, modified by labeling characteristic geological sites. The water-rich plumes erupt from the warm fractures near the South Pole region. The plumes originate from the global subsurface ocean that interacts with the rocky core. On the seafloor or within the rocky core, there are alkaline hydrothermal activities at temperatures $>90^\circ\text{C}$, generating H_2 and nanosilica particles (Image courtesy: NASA/JPL-Caltech (PIA 19058))

25.2 Liquid Water on Enceladus

Multiple lines of evidence demonstrate the presence of a subsurface ocean within Enceladus (e.g., Schmidt et al. 2008; Waite Jr et al. 2009; Postberg et al. 2009, 2011). The “smoking gun” for the presence of the subsurface ocean is the highly saline (0.5–2% of NaCl relative to H₂O) icy particles in Enceladus’ plume (Postberg et al. 2009, 2011). The measured high salinity is not possible merely by sublimation of the icy crust, but it can be achieved only by dissolution of Na⁺ and Cl⁻ ions from the rocky core into the subsurface ocean via water-rock interactions (Zolotov 2007, 2012; Postberg et al. 2009).

The presence of liquid water is also supported by a series of the observations of plume dynamics, i.e., the difference in velocity between ice particles and vapor of the plume, ice-to-water ratio of the plume materials, and strong infrared radiation (~4.2 billion watts or GW) near the tiger stripes—four fractures bounded by the ridges near the South Pole region (e.g., Porco et al. 2006; Spencer et al. 2013). Numerical models, supporting these observations, suggest the occurrence of upward ejection of water-rich vapor and liquid droplets at the interface between the subsurface ocean and icy crust near the triple point of H₂O (i.e., ~270 K) (e.g., Schmidt et al. 2008). Then, partial condensation of water vapor occurs within the cracks of the icy crust, releasing latent heat to the crust and accelerating remaining vapor gas toward the surface (e.g., Ingersoll and Pankine 2010; Nakajima and Ingersoll 2016). Collisions of the particles with the wall of the cracks also result in deceleration of icy particles (e.g., Schmidt et al. 2008). Given the presence of antifreeze compounds, such as NH₃ and CH₃OH, in the plume (Waite Jr et al. 2009), the constraint for the temperature of the plume source (i.e., ~270 K; Schmidt et al. 2008) strongly supports the presence of liquid water within Enceladus.

Recent gravity, shape, and libration data provided by the Cassini spacecraft suggest that Enceladus’ ocean is most likely global (Fig. 25.1) (e.g., Iess et al. 2014; McKinnon 2015; Beuthe et al. 2016; Čadek et al. 2016; Thomas et al. 2016; Van Hoolst et al. 2016). These studies infer that the icy crust has a mean thickness of ~20 to 30 km and a thickness beneath the South Pole region of less than several km (viz., a mean thickness of the subsurface ocean of 30–40 km and that for the South Pole region of >60 km) (Beuthe et al. 2016; Čadek et al. 2016; Thomas et al. 2016; Van Hoolst et al. 2016). The thin icy crust (i.e., a few km in thickness) near the South Pole is also supported by the observation of thermally anomalous features revealed by Cassini’s RADAR instrument (Le Gall et al. 2017). To sustain the global ocean within Enceladus, anomalously high levels of heat production rate are required. In a scenario where maximum heat is produced in the rocky core, it needs to be approximately ten times that of radiogenic heating expected for Enceladus’ rocky core (Kamata and Nimmo 2017). In another scenario where maximum heat is produced in the icy crust via equilibrium tides, it also needs to be about ten times the conventional estimate (Kamata and Nimmo 2017). The recent study suggests that a porous, unconsolidated rocky core with a low effective rigidity can generate high levels of heat by tidal dissipation with the core (Choblet et al. 2017). Coupled numerical simulations of tidal friction and water transfer within the porous, unconsolidated

core suggest that more than 10 GW of heat can be generated inside the core, which can sustain the global ocean (Choblet et al. 2017). In addition, they show that high-temperature fluids would transfer in narrow regions within the core, which favors high-temperature water-rock interactions (Choblet et al. 2017). Nevertheless, their model assumes the core with an extremely low effective rigidity and high dissipation factor. The validity of this assumption is unknown.

The major unsolved question is how long the high levels of power have been generated within Enceladus. This largely depends on how the orbital eccentricity of Enceladus evolves throughout its history. Stored tidal heat released episodically could explain the high heat production rate (e.g., Meyer and Wisdom 2007; O'Neill and Nimmo 2010; Shoji et al. 2014). Another model suggests that a large quantity of tidal heat (e.g., several tens of GW) is generated within the icy crust of Enceladus via dynamical tides (Fuller et al. 2016). If the latter is the case, the tidal heat might sustain the global ocean and hydrothermal activity for a long time period (e.g., billions of years) within Enceladus (Choblet et al. 2017).

Another novel hypothesis suggests that Enceladus and the other inner satellites of Saturn would have been formed only in ~ 100 Myrs ago (Ćuk et al. 2016), from a massive ring of the gas giant by subsequent outward migration (Crida and Charnoz 2012). This model is completely divergent from the earlier proposed formation models of icy satellites within a circumplanetary disk (e.g., Canup and Ward 2006; Sekine and Genda 2012). If this model is corroborated, Enceladus and the other mid-sized satellites would not necessarily be ~ 4.5 Gyrs old, and its high heat production rate could be explained by accretion remnants and/or tidal heating due to the recent orbital evolution (Ćuk et al. 2016). However, one major inconsistency with the recent formation model is the heavily cratered and ancient surface of Mimas (~ 4 Gyrs old according to crater chronology) (Jaumann et al. 2009), the innermost regular satellite of Saturn.

How old is the ocean on Enceladus? What is the mechanism that sustains the high heat production? And, how long will it sustain in the future? The sustained oceanic lifetime and heat production rates are two vital factors for the chemical evolution of simple organics to form complex organic matter and subsequently the origin of life within Enceladus (e.g., McKay et al. 2008).

25.3 Organic Matter

The INMS and CDA of the Cassini spacecraft detected a variety of organic compounds in Enceladus' plume (e.g., Waite Jr et al. 2009). Methane, CH_4 , the most abundant aliphatic hydrocarbon in the plume with concentration of 0.1–1% relative to H_2O , and NH_3 , the most N-abundant compound in the plume with concentration of 0.1–2% relative to H_2O , were identified by the INMS (Waite Jr et al. 2009; Waite et al. 2017). Furthermore, the INMS also possibly detected the ion masses corresponding to higher aliphatic and aromatic C_2 – C_6 hydrocarbons, like C_2H_6 and C_6H_6 ($\sim 10^{-3}$ – 10^{-1} % relative to H_2O), and hydrogen cyanide (HCN) in the gas-phase

compound of the plume (e.g., Waite Jr et al. 2009). Cassini's UVIS showed that gaseous N_2 would be much less than 0.5% relative to H_2O (Hansen et al. 2011).

The solid-phase component of the plume was also found to contain high-molecular-weight organic matter with largely unknown chemical composition and structure, which indicates its presence in the subsurface ocean (Postberg et al. 2011, 2018). According to the detailed analysis for the CDA mass spectra, the molecular mass of the organic matter exceeds 200 atomic mass units (Postberg et al. 2018). The results from Cassini's CDA also imply the presence of unsaturated aromatic hydrocarbons and amine functional group ($-NH_2$) in the high-molecular-weight organic matter (Khawaja et al. 2015; Postberg et al. 2018).

Not much is known about the origin of this high-molecular-weight organic matter, namely, whether it is synthesized by the geochemical processes occurring on Enceladus and whether it represents the starting building materials of this moon. The satellite formation models suggest that the chemical compositions of building materials of Enceladus would have been similar to those of comets (e.g., Canup and Ward 2006), of exogenic origin, and containing huge quantities of high-molecular-weight organic matter (e.g., Keller et al. 2006).

The alternate geochemical synthesis hypothesis proposes that the high-molecular-weight organic matter in the plume may have been synthesized by polymerization of simple C- and N-bearing compounds, such as HCHO, HCN, and NH_3 , under hydrothermal conditions within Enceladus. These simple molecules are commonly found in comets (e.g., Bockelée-Morvan et al. 2004; Goesmann et al. 2015) and are known to polymerize in an alkaline aqueous environments at high temperatures (e.g., 100 °C), to form high-molecular-weight organic matter (Cody et al. 2011; Kebukawa et al. 2013; Sekine et al. 2017). The pH and temperature of Enceladus' hydrothermal system (pH ~9 to 11 and temperature of >90 °C; Hsu et al. 2015; Sekine et al. 2015) are suitable for polymerization of these C- and N-bearing molecules. The future space missions bound for Enceladus should investigate whether prebiotic organic molecules, including those that are functional and informational biopolymer of life (e.g., proteins, DNA, and RNA for Earth's life), are generated within Enceladus' ocean (e.g., McKay et al. 2008).

25.4 Bioavailable Energy

Life requires chemical energy to support the synthesis of functional and informational organic materials. The chemical affinity that can be used by the chemosynthetic microorganisms on Earth is produced and sustained through the chemical reactions between reductants and oxidants (e.g., McCollom and Shock 1997; Amend et al. 2011). In these settings on Earth, the water-rock reactions in hydrothermal systems on the seafloor provides reductants, such as H_2 (e.g., Yoshizaki et al. 2009; Mayhew et al. 2013); whereas, oxidants, such as CO_2 and SO_4^{2-} , are provided by volcanoes and atmospheric processes. Since Enceladus' ocean is subsurface, it does not receive any incoming solar energy. Accordingly, chemical redox disequilibrium

could be a major energy source for life on Enceladus. To sustain such redox chemical processes and life, nevertheless, there is a need to identify zones with hydrothermal activity and to detect reductants and oxidants in Enceladus' ocean.

The presence of nanosilica particles contained in Saturn's E-ring is a crucial evidence for an ongoing hydrothermal activity within Enceladus (Hsu et al. 2015). Sekine et al. (2015) further have shown that the rock component of the hydrothermal activity is most probably chondrite-like. In high-temperature, water-rock interactions, SiO_2 from the rock component dissolves in hydrothermal fluids, which further upon cooling condenses to form nanosilica particles. Laboratory simulation experiments of hydrothermal reactions have constrained the required temperatures to >90 °C for formation of nanosilica in Enceladus (Fig. 25.1) (Hsu et al. 2015; Sekine et al. 2015). Given the nanoscale size of these silica particles, Hsu et al. (2015) also have suggested its residence time in the subsurface ocean within several Earth's years, thus indicating an ongoing alkaline hydrothermal activity.

One well-known terrestrial analogues of such an alkaline, moderate-temperature hydrothermal system hosted by ultramafic rocks is the Lost City Hydrothermal Field in the Atlantic Ocean (e.g., Kelley et al. 2005). This hydrothermal field supports a variety of chemosynthetic microorganisms potentially including hydrogenotrophic methanogens (e.g., Kelley et al. 2005). The hydrothermal systems in the Hadean-Archean eons (4.5–2.5 Gyrs ago) on Earth hosted by mafic-ultramafic rocks probably resulted in a CO_2 - and H_2 -rich mixing zone between seawater and hydrothermal vent fluid (Shibuya et al. 2010, 2013, 2015; Ueda et al. 2016). Such mixing zone could generate sufficient bioavailable energy for hydrogenotrophic methanogens (Shibuya et al. 2016). The Hadean-Archean hydrothermal systems (Shibuya et al. 2010) could have been analogous to Enceladus's hydrothermal environments.

Hydrothermal reactions within Enceladus' chondritic core would generate a large quantity of H_2 through oxidation of Fe(metal) and/or Fe(II) by H_2O (Glein et al. 2015; Sekine et al. 2015). In fact, recent observations of the gas-phase component of Enceladus' plume revealed the presence of $\sim 1\%$ of H_2 relative to H_2O (Waite et al. 2017). As the plume contains $\sim 1\%$ of CO_2 relative to H_2O (Waite et al. 2017), coexistence of H_2 and CO_2 is suitable for methanogenic life (McKay et al. 2008). Waite et al. (2017) calculated the chemical affinity of CH_4 formation within Enceladus' ocean from the measured amounts of H_2 and CO_2 as 40–100 kJ/mol CH_4 , which is higher than chemical energy required for ATP synthesis (~ 20 kJ/mol).

Although the present-day Enceladus is possibly energetically habitable, it is unknown if the CH_4 cycles back to CO_2 and H_2 in geochemical cycles to sustain chemical disequilibrium for longer geological timescales on Enceladus. Solar UV-based photolysis occurring on the surface of the icy crust could promote dissociation of CH_4 and H_2O , possibly leading to generation of CO_2 . If geological recycling of the icy crust occurs on Enceladus, this process could be a source of CO_2 .

Chemical disequilibrium can also occur through pyrolysis of CH_4 and H_2O within the hot rocky core. Thermochemical calculations show that CH_4 and H_2O are unstable at >200 °C, compared with CO_2 and H_2 (Glein et al. 2008). However, it is highly uncertain whether such high temperatures can be achieved within Enceladus given its small size. Additionally, even if high-temperature hydrothermal systems occur within

Enceladus, the reaction kinetics of the thermal decomposition of CH_4 could prevent the recycling of CH_4 . If the recycling of bioavailable CH_4 is inefficient in geochemical cycles within Enceladus, disequilibrium energy for life would only be temporary.

25.5 Summary

Owing to Cassini's observations, the present-day Enceladus is now known to possess a habitable environment, where liquid water, organic matter, and bioavailable energy coexist. Yet it is not known if these three factors can be maintained over long geological timescales on Enceladus. If sustained for sufficiently long duration, however, compared with that for the emergence of life on Earth (e.g., >0.5 Gyrs), Enceladus would become an important site to search for signs of life beyond Earth. The detection or non-detection of life on Enceladus would provide insights into the elemental/geochemical requirements for life. Comparisons between Enceladus and early Earth, in terms of the roles of surface oceans, a substantial atmosphere, and lands for the origin of life, could be studied. Even if the habitable environment on Enceladus is short-lived (e.g., ~0.1 Gyrs or less), it would enhance our understanding about the chemical evolution toward life, which cannot be achieved merely by our current knowledge. Whatever the case, Enceladus is a unique planetary body from the standpoint of astrobiological research.

Acknowledgments This work was supported by MEXT KAKENHI Grant Number JP 17H0655, 17H06456, and 17H06457.

References

- Amend JP, McCollom TM, Hentscher M, Bach W (2011) Catabolic and anabolic energy for chemolithoautotrophs in deep-sea hydrothermal systems hosted in different rock types. *Geochim Cosmochim Acta* 75:5736–5748
- Beuthe M, Rivoldini A, Trinh A (2016) Enceladus's and Dione's floating ice shells supported by minimum stress isostasy. *Geophys Res Lett* 43:10088–10096
- Bockelée-Morvan D, Crovisier J, Mumma MJ, Weaver HA (2004) The composition of cometary volatiles. In: Festou MC, Keller HU, Weaver HA (eds) *Comets II*. Univ. Arizona Press, Tucson, pp 391–423
- Čadež O et al (2016) Enceladus's internal ocean and ice shell constrained from Cassini gravity, shape and libration data. *Geophys Res Lett* 43:5653–5660
- Canup RM, Ward WR (2006) A common mass scaling for satellite systems of gaseous planets. *Nature* 441:834–839
- Choblet G et al (2017) Powering prolonged hydrothermal activity inside Enceladus. *Nature Astron* 1:841–847. <https://doi.org/10.1038/s41550-017-0289-8>
- Cody GD, Heying E, Alexander CMO, Nittler LR, Kilcoyne ALD, Sandford SA, Stroud RM (2011) Establishing a molecular relationship between chondritic and cometary organic solids. *Proc Natl Acad Sci* 108:19171–19176

- Crida A, Charnoz S (2012) Formation of regular satellites from ancient massive rings in the solar system. *Science* 338:1196–1199
- Čuk M, Dones L, Nesvorný D (2016) Dynamical evidence for a late formation of Saturn's moons. *Astrophys J* 820:97 (16 pp)
- Fuller J, Luan J, Quataert E (2016) Resonance locking as the source of rapid tidal migration in the Jupiter and Saturn moon systems. *Mon Not R Astron Soc* 458:3867–3879
- Glein CR, Zolotov MY, Shock EL (2008) The oxidation state of hydrothermal systems on early Enceladus. *Icarus* 197:157–163
- Glein CR, Baross JA, Waite JH Jr (2015) The pH of Enceladus' ocean. *Geochim Cosmochim Acta* 162:202–219
- Goesmann F et al (2015) Organic compounds on comet 67P/Churyumov-Gerasimenko revealed by COSAC mass spectrometry. *Science* 349:aab0689 1–3
- Hansen CJ et al (2011) The composition and structure of the Enceladus plume. *Geophys Res Lett* 38:L11202. <https://doi.org/10.1029/2011GL047415>
- Hsu H-W et al (2015) Silica nanoparticles as an evidence of hydrothermal activities at Enceladus. *Nature* 519:207–210
- Iess L et al (2014) The gravity field and interior structure of Enceladus. *Science* 344:78–80
- Ingersoll AP, Pankine AA (2010) Subsurface heat transfer on Enceladus: conditions under which melting occurs. *Icarus* 206:594–607
- Jaumann R et al (2009) Icy satellites: geological evolution and surface processes. In: Dougherty M, Esposito L, Krimigis S (eds) *Saturn from Cassini-Huygens*. Springer, Heidelberg, pp 637–681
- Kamata S, Nimmo F (2017) Interior thermal state of Enceladus inferred from the viscoelastic state of the ice shell. *Icarus* 284:387–393
- Kebukawa Y, Kilcoyne ALD, Cody GD (2013) Exploring the potential formation of organic solids in chondrites and comets through polymerization of interstellar formaldehyde. *Astrophys J* 771(19):1–12
- Keller LP et al (2006) Infrared spectroscopy of comet 81P/wild 2 samples return by stardust. *Science* 314:1728–1731
- Kelley DS et al (2005) A serpentine-hosted ecosystem: the lost city hydrothermal field. *Science* 307:1428–1434
- Khawaja N et al (2015) Organic compounds from Enceladus' sub-surface ocean as seen by CDA. In: *European Planetary Science Congress 2015*, 10: 652
- Le Gall A et al (2017) Thermally anomalous features in the subsurface of Enceladus's south polar terrain. *Nature Astron* 1:0063. <https://doi.org/10.1038/s41550-017-0063>
- Mayhew LE, Ellison ET, McCollom TM, Trainor TP, Templeton AS (2013) Hydrogen generation from low-temperature water-rock reactions. *Nat Geosci* 6:478–484
- McCollom TM, Shock EL (1997) Geochemical constraints on chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal systems. *Geochim Cosmochim Acta* 61:4375–4391
- McKay CP, Porco CC, Altheide T, Davis WL, Kral TA (2008) The possible origin and persistence of life on Enceladus and detection of biomarkers in the plume. *Astrobiology* 8:909–919
- McKinnon WB (2015) Effect of Enceladus's rapid synchronous spin on interpretation of Cassini gravity. *Geophys Res Lett* 41:2137–2143
- Meyer J, Wisdom J (2007) Tidal heating in Enceladus. *Icarus* 188:535–539
- Nakajima M, Ingersoll AP (2016) Controlled boiling on Enceladus. 1. Model of the vapor-driven jets. *Icarus* 272:309–318
- O'Neill CO, Nimmo F (2010) The role of episodic overturn in generating the surface geology and heat flow on Enceladus. *Nat Geosci* 3:88–91
- Porco CC et al (2006) Cassini observes the active South Pole of Enceladus. *Science* 311:1393–1401
- Postberg F et al (2009) Sodium salts in E-ring ice grains from an ocean below the surface of Enceladus. *Nature* 459:1098–1101
- Postberg F, Schmidt J, Hillier J, Kempf S, Srama R (2011) A salt-water reservoir as the source of a compositionally stratified plume on Enceladus. *Nature* 474:620–622

- Postberg F et al (2018) Macromolecular organic compounds from the depths of Enceladus. *Nature* 558:564–568
- Schmidt J, Brilliantov N, Spahn F, Kempf S (2008) Slow dust in Enceladus' plume from condensation and wall collisions in tiger stripe fractures. *Nature* 451:685–688
- Sekine Y, Genda H (2012) Giant impacts in the Saturnian system: a possible origin of diversity in the inner mid-sized satellites. *Planet Space Sci* 63–64:133–138
- Sekine Y et al (2015) High-temperature water-rock interactions and hydrothermal environments in the chondrite-like core of Enceladus. *Nature Comm* 6:8604. <https://doi.org/10.1038/ncomms9604>
- Sekine Y, Genda H, Kamata S, Funatsu T (2017) The Charon-forming giant impact as a source of Pluto's dark equatorial regions. *Nature Astron* 1:0031. <https://doi.org/10.1038/s41550-016-0031>
- Shibuya T, Komiya T, Nakamura K, Takai K, Maruyama S (2010) Highly alkaline, high-temperature hydrothermal fluids in the early Archean ocean. *Precambrian Res* 182:230–238
- Shibuya T et al (2013) Reactions between basalt and CO₂-rich seawater at 250 and 350°C, 500 bars: implications for the CO₂ sequestration into the modern oceanic crust and composition of hydrothermal vent fluid in the CO₂-rich early ocean. *Chem Geol* 359:1–9
- Shibuya T et al (2015) Hydrogen-rich hydrothermal environments in the Hadean ocean inferred from serpentinization of komatiites at 300 °C and 500 bar. *Prog Earth and Planet Sci* 2:46. <https://doi.org/10.1186/s40645-015-0076-z>
- Shibuya T, Russell M, Takai K (2016) Free energy distribution and chimney minerals in Hadean submarine alkaline vent systems; importance of iron redox reactions under anoxic condition. *Geochim Cosmochim Acta* 175:1–19
- Shoji D, Hussmann H, Sohl F, Kurita K (2014) Non-steady state tidal heating of Enceladus. *Icarus* 235:75–85
- Spencer JR et al (2013) Enceladus heat flow from high spatial resolution thermal emission observations. In: *European Planetary Space Congress 2013*, 8: 840
- Thomas R et al (2016) Enceladus's measured physical libration requires a global subsurface ocean. *Icarus* 264:37–47
- Ueda H et al (2016) Reactions between komatiite and CO₂-rich seawater at 250 °C and 350 °C, 500 bars: implications for hydrogen generation in the Hadean seafloor hydrothermal system. *Prog Earth Planet Sci* 3:35. <https://doi.org/10.1186/s40645-016-0111-8>
- Van Hoolst T, Baland R–M, Trinh A (2016) The diurnal libration and interior structure of Enceladus. *Icarus* 277:311–318
- Waite JH Jr et al (2009) Liquid water on Enceladus from observations of ammonia and ⁴⁰Ar in the plume. *Nature* 460:487–490
- Waite JH et al (2017) Cassini finds molecular hydrogen in the Enceladus plume: evidence for hydrothermal processes. *Science* 356:155–159
- Yoshizaki M et al (2009) H₂ generation by experimental hydrothermal alteration of komatiitic glass at 300°C and 500 bars: a preliminary result from on-going experiment. *Geochem J* 43:17–22
- Zolotov MY (2007) An oceanic composition on early and today's Enceladus. *Geophys Res Lett* 34:L23203. <https://doi.org/10.1029/2007GL031234>
- Zolotov MY (2012) Aqueous fluid composition in CI chondritic materials: chemical equilibrium assessments in closed systems. *Icarus* 220:713–729

Chapter 26

Astrobiology on Titan: Geophysics to Organic Chemistry



Hiroshi Imanaka

Abstract Titan, the largest satellite of Saturn, is the only moon with a substantial atmosphere in our solar system. The Cassini-Huygens mission by NASA/ESA (2004–2017) returned a wealth of information about Titan’s atmosphere and surface environments. Titan exhibits remarkable similarities with Earth. Active organic chemistry in the atmosphere, a dynamic methane hydrological cycle, and the internal global water ocean are all unique features of astrobiological interest. The exploration of Titan’s organic environment provides us with a unique opportunity to understand abiotic, possibly prebiotic, chemistry on a planetary scale.

Keywords Titan · Organic chemistry · Hydrological cycle · Prebiotic chemistry · Habitable environment

26.1 Titan: Interests for Astrobiology

Titan, the largest satellite of Saturn, is the only solar system moon with a substantial atmosphere. Its atmosphere is composed primarily of N_2 with a few percent of CH_4 . The surface pressure is 1.5 bar with a surface temperature of 94 K. The mean density of Titan is 1.88 g/cm^3 indicating a bulk composition of roughly equal mixture of rock and water ice. The Voyager observations in the early 1980s revealed the globally covering thick haze layers, completely obscuring the surface at visible wavelengths. Several simple hydrocarbon and nitrile species identified in the stratosphere and the similarity in the orange-reddish color of organic solids generated in a N_2 - CH_4 plasma discharge experiments (termed “Titan tholin” by Sagan and Khare) lead to the hypothesis of ongoing complex organic chemistry in Titan’s atmosphere (Sagan et al. 1984).

H. Imanaka (✉)
NASA Ames Research Center, Moffett Field, CA, USA
SETI Institute, Mountain View, CA, USA
e-mail: himanaka@seti.org

Titan is often considered as one of the best targets to study prebiotic chemistry at a full planetary scale, even though Titan's surface temperature of 94 K is too cold for liquid water to be present. Titan's mildly reduced N₂ atmosphere and the rich organic chemistry may share many similarities to the primitive Earth atmosphere before life arose. Since the geological records of the early Earth environment when life originated have been essentially destroyed, the exploration of Titan's organic environment provides us a unique opportunity to understand abiotic, possibly prebiotic, chemistry on a planetary scale.

The Cassini-Huygens mission (NASA/ESA 2004–2017) has revolutionized our view of Titan (Fig. 26.1) as described in the next section. In fact, Titan shares remarkable similarities with the Earth. It features the most complex organic chemistry known outside of Earth and, uniquely, hosts an analog to Earth's hydrological cycle, with methane-forming clouds, rain, rivers, and lakes/seas of methane. It also exhibits various fluvial/erosional/aeolian geological units similar to Earth.

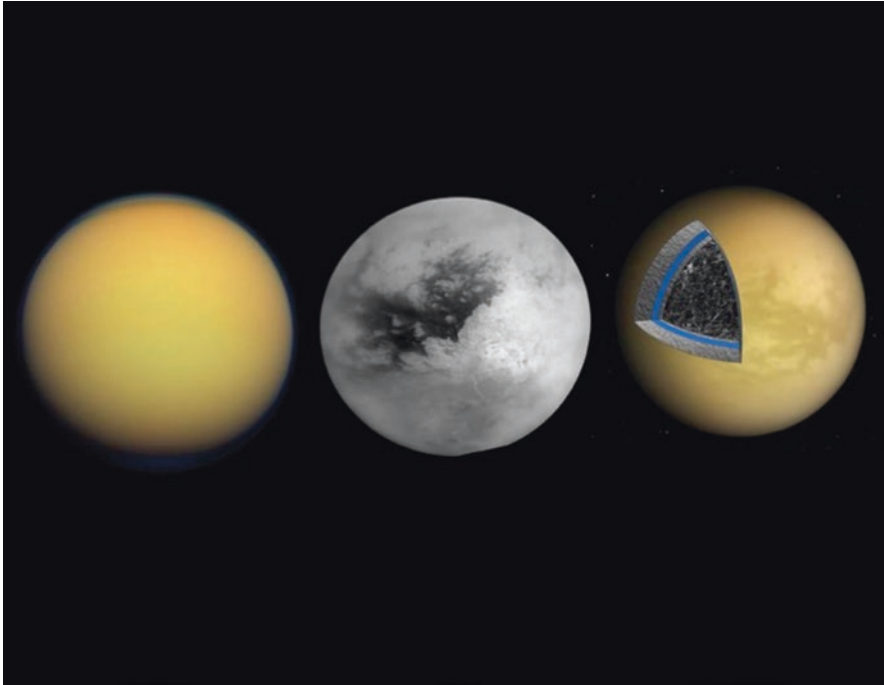


Fig. 26.1 Titan's atmosphere, surface features, and interior structure. (a) A natural color image by the Cassini Imaging Science Subsystem (ISS) (PIA14602, NASA/JPL-Caltech/Space Science Institute) (Porco et al. 2005). Globally covered haze layers obscuring the surface. (b) Surface features at the near-infrared wavelength (ISS) revealed the presence of geologically distinct surface areas whose exact nature and composition remains largely unknown (PIA06185, NASA/JPL-Caltech/Space Science Institute). (c) An interior model inferred from Cassini's gravity observations. Subsurface global ocean, with high pressure phase

Furthermore, multiple evidences suggest the presence of a subsurface global water ocean. Thus, Titan has an active and dynamical environment very rich in organics, circulating and probably evolving on a planetary scale. It might even possess conditions (organics, water, and energy) for life to originate. Thus, Titan becomes an even more fascinating target for astrobiology than ever previously believed for this body.

In this chapter, we briefly summarize the astrobiological aspects of Titan and its potential regarding future missions. Greater details should be referred to the comprehensive books on Titan (Brown et al. 2009; Müller-Wodarg et al. 2014) and review articles (e.g., Raulin et al. 2012; Mitchell and Lora 2016; Hayes 2016; Horst 2017).

26.2 Cassini-Huygens Mission (2004–2017)

The Cassini-Huygens mission by NASA/ESA (2004–2017) returned a wealth of information regarding Titan's atmosphere and surface environments. The main constituents of Titan's atmosphere are N_2 and a few percent of CH_4 . Solar radiation and Saturn's magnetospheric charged particles drive active organic chemistry in Titan's atmosphere (Fig. 26.2). One of the most surprising results from the Ion Neutral Mass Spectrometer (INMS) and the Cassini Plasma Spectrometer (CAPS) is the

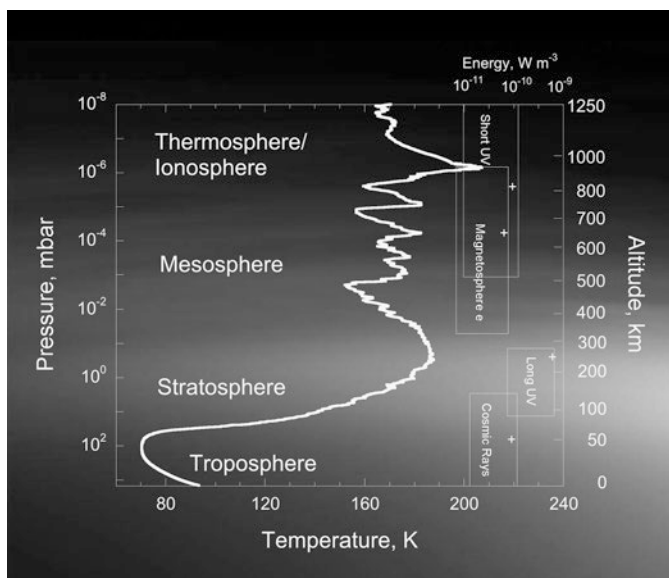


Fig. 26.2 Atmospheric structure and energy sources. Titan's thermal structure measured by the Huygens Atmospheric Structure Instrument (HASI) (Fulchignoni et al. 2005), overlaid (and approximately scaled) to the Cassini ISS image showing thick cloud deck and multiple detached haze layers (PIA06236, NASA/JPL/Space Science Institute). The major energy sources deposited in the Titan atmosphere are also shown. (Updated from Imanaka et al. 2004)

formation of complex organic molecular ions over 3000 atomic mass units in the ionosphere of Titan (Waite et al. 2007). The Cassini Ultraviolet Imaging Spectrometer (UVIS) observation reveals aerosol particles widely distributed in Titan's thermosphere/mesosphere (Koskinen et al. 2011). This observation clearly demonstrated the importance of complex organic chemistry in the upper atmosphere induced by extreme ultraviolet (EUV) photons and Saturn's magnetospheric charged particles (Lavvas et al. 2013). However, the exact chemical nature of the haze particles is still unknown though the spectral features obtained from the Visible Infrared Mapping Spectrometer (VIMS) and Composite Infrared Spectrometer (CIRS) are indicative of a high abundance of hydrocarbon species (Imanaka et al. 2012).

In the stratosphere, the CIRS instrument observed emission bands from methane and numerous higher-order hydrocarbons (C_2H_2 , C_2H_4 , C_2H_6 , C_3H_8 , CH_3C_2H , C_4H_2), several nitriles (HCN, HC_3N , and C_2N_2), and oxygen containing species (CO, CO_2 , and H_2O) (e.g., Bezard 2009). Their vertical and latitudinal distributions and their seasonal variations revealed the interplay among chemistry, radiation, and dynamics in the atmosphere (e.g., Mitchell and Lora 2016). The spectral evidence of condensation clouds is also observed, but the exact chemical nature is still not well understood (Anderson and Samuelson 2011). The GCMS on the Huygens probe sampled gaseous species through the descent in the Titan's stratosphere and troposphere (Niemann et al. 2005). Lack of volatile species in the troposphere except N_2 , CH_4 , and H_2 is consistent with condensations of other volatile species high in the stratosphere. Aerosol samples collected by the Aerosol Collector and Pyrolyzer (ACP) revealed the presence of NH_3 and HCN upon pyrolysis at 600 C, which was interpreted as the evidence of nitrogenated complex organic aerosols (Israel et al. 2005). However, Biemann (2006) argued against the identification and interpretation of NH_3 and HCN. After landing on Titan's surface, the GCMS detected an increased level of CH_4 vapor, indicative of a moist nature to the surface material. Surface spectra taken by the Descent Imager/Spectral Radiometer (DISR) show a featureless blue slope between 800 and 1500 nm that matches no mixture of laboratory spectra of pure ices, organics, or tholins (Tomasko et al. 2005).

The organics appear to be ubiquitous and compositionally dominant on Titan's surface (e.g., Soderblom et al. 2007; Barnes et al. 2009). No pure water ice has been identified on the surface. The bright regions in the Imaging Science Subsystem (ISS) near-infrared image are consistent with organics, while dark regions as contaminated water ice (Fig. 26.1b). However, the exact determination of the chemical composition of Titan's surface suffers from the opaque atmosphere limiting the remote sensing observations to a few spectral regions. The Cassini RADAR revealed surface material of low dielectric constant (Janssen et al. 2016). This is consistent with organic materials of a ~ 1 m depth above a bedrock of water ice, implying the accumulation of atmospheric organic haze on the surface.

The Cassini orbiter and the Huygens probe revealed an Earthlike landscape of fluvial valleys, channels, lakes, and extensive dune fields (Fig. 26.3), indicating an active dynamic hydrological cycle on Titan (e.g., Lunine and Atreya 2008). The RADAR data from the north polar region revealed convincing evidence of lakes composed of liquid methane and ethane (Stofan et al. 2007; Hayes 2016) (Fig. 26.3a). The

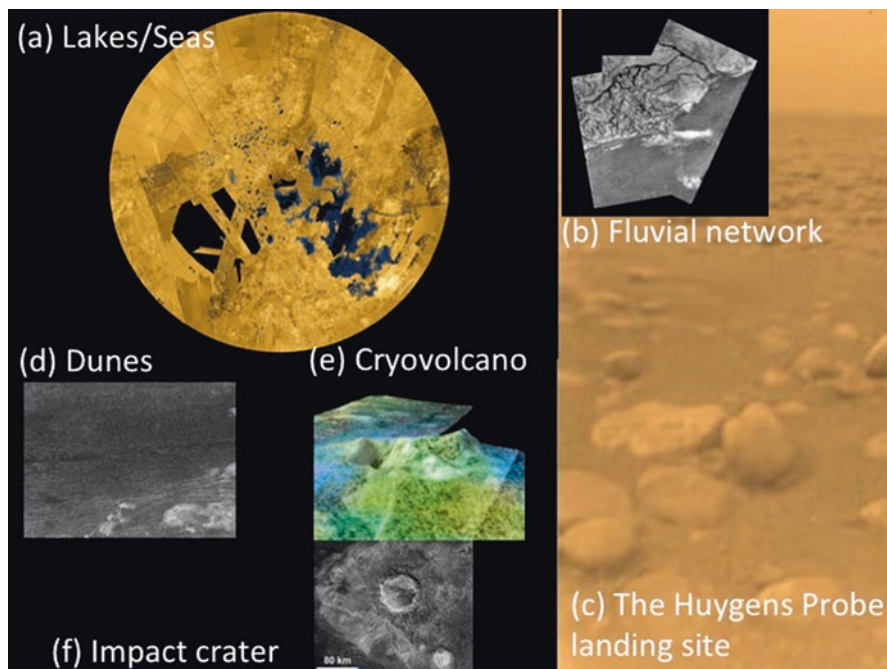


Fig. 26.3 Titan's geological features. **(a)** Widespread presence of lakes/seas in Titan's northern high latitudes revealed by the RADAR (PIA17655, NASA/JPL-Caltech/ASI/USGS). **(b)** Dendritic channel system as evidence of fluvial activity near the Huygens landing site (PIA07236, NASA/JPL/ESA/University of Arizona). **(c)** The images taken after the Huygens Probe landing shows the rounded pebbles of 10–15 cm, strongly supporting fluvial activity (PIA07232, NASA/JPL/ESA/University of Arizona). **(d)** Vast fields of dunes around Titan's equatorial region, composed of millimeter to submillimeter particles driven by winds (PIA08738, NASA/JPL-Caltech/ASI). **(e)** Possible cryovolcano, Sotra Patera, a circular mountain structure with its computer 3-D reconstruction (NASA/JPL-Caltech/ASI/USGS/University of Arizona). **(f)** A relatively fresh crater called Sinlap (PIA16638, NASA/JPL-Caltech/ASI/GSFC)

equatorial dark regions in ISS images have been shown to consist of extensive fields of dry dunes (Lorenz et al. 2006). The optically dark appearance of the dune material and its spectral characteristics support the presence of surface organic compounds, which represent the largest known reservoir of organic material on Titan (Lorenz et al. 2008). The Huygens probe images of dendritic channel networks, shorelines, and rounded pebbles on Titan convincingly show that Titan, like Earth, has a surface carved by flowing fluids through active hydrological cycle (Tomasko et al. 2005).

The endogenic and exogenic geological features are of particular interest in terms of exchanging materials from the interior, such as exposing water-ammonia mixtures on the surface. The paucity of impact features indicates that the surface of Titan is very young, an age less than 1 Gyr. Probable cryovolcanic features at Sotra Patera (Fig. 26.3e) are suggested (Lopes et al. 2013). Cryovolcanism can be a mechanism to replenish CH_4 to the atmosphere; otherwise the CH_4 in the

atmosphere would be totally photolyzed within 10–20 Myr (Lunine and Atreya 2008). However, the evidence for cryovolcanic features on Titan is still on debate.

Deep beneath the frigid surface, the Cassini orbiter and Huygens probe found evidence for a liquid water ocean (Fig. 26.1c), probably ammonia-rich, serving as an antifreezing agent (e.g., Tobie et al. 2006). Electric signals measured in Titan's atmosphere by the Huygens probe suggest the presence of a conductive layer some 55–80 km below Titan's surface, possibly a water ocean doped with small amounts of salts or ammonia to increase the electrical conductivity (Béghin et al. 2009). Titan's spin state and tidal gravity response are also suggestive of Titan's ice crust being decoupled from the deep interior, hence by a global internal ocean layer, roughly 100 km below the surface (Iess et al. 2012).

26.3 Prebiotic-Like Chemistry and Habitability on Titan

Titan's atmosphere-surface environment may provide great observational tests of various hypotheses for the origin of life on Earth. All the ingredients, which are presumed necessary for life – liquid-water, organic matter, and energy – seem present on Titan. Active organic chemistry in Titan N_2 - CH_4 atmosphere generates a wide variety of organic species in volatiles, condensates, and refractory materials. Those organic species are eventually accumulated on the surface of Titan, transferred by the active hydrological cycle. High-energy cosmic ray or meteor impacts can drive further complex chemistry on the Titan surface. Titan clearly provides a variety of geologic environments for the staging of organic chemistry, in which there exists free energy for reactions. Complex organic materials of a wide variety of reactive species could undergo further chemical evolution.

Even though the Cassini-Huygens mission has put tighter constraints on Titan's atmospheric chemistry and complex organic haze, the exact chemical composition of the haze particles is still a mystery. Thus, our understanding of the nature of this haze has been mainly built upon laboratory and theoretical models. Laboratory experiments simulating Titan's haze material (“tholins”) provide a key source of understanding of the nature of the Titan haze and the processes that form it (e.g., see review in Cable et al. 2011). Though significant progress regarding nascent tholin molecular composition has been made (e.g., Imanaka and Smith 2010; Cable et al. 2011), a complete tholin inventory eludes determination after several decades of study. Investigation of prebiotic molecular formation upon hydrolysis of tholins or gas phase chemistry has provided insights into the chemical potential of prebiotic evolution on Titan (e.g., Cable et al. 2011; Raulin et al. 2012).

Although Titan's surface is too cold to support stable liquid water, temporal pools of water/ammonia solutions could have existed on the surface created by impact or cryovolcanism. The VIMS spectral properties of Titan's impact craters suggest the exposure of an intimate mixture of water ice and organic materials (Neish et al. 2015). Interactions between reactive organic molecules and liquid water could initiate further chemical processes, possibly prebiotic-like evolution for

modest periods on Titan (Raulin et al. 2012; Lunine 2017). A subsurface ocean of liquid water-ammonia could also be a tempting candidate for a possible habitable environment, since it might contain both the organic materials and the energy sources necessary for life. At the beginning of Titan's history, a global ocean could have been in direct contact with the atmosphere and with the internal rocky core, offering interesting analogies with the primitive Earth and the potential implication of hydrothermal vents in terrestrial prebiotic chemistry (Lunine 2017).

There have been some speculations of a completely different form of life to exist in the liquid hydrocarbon lakes on Titan. Benner et al. (2004) suggested that the liquid hydrocarbons on Titan could be the basis for life, playing the role that water does for life on Earth. Stevenson et al. (2015) proposed a possible membrane formation in liquid methane, which could serve as cell boundary. A hypothetical second form of life independent of the water-based life we know on the Earth could use H_2 and C_2H_2 to derive free energy (McKay and Smith 2005). In fact, a disparity in the hydrogen densities that lead to a flow down to the surface was inferred to explain the vertical distribution of chemical species (Strobel 2010). A non-biological origin of this phenomenon has not been ruled out; however, it is worthwhile keeping our eyes widely open to life mechanisms we don't know of. McKay (2016) summarized an approach to characterizing Titan as a possible abode of life.

26.4 Future Titan Explorations

The Cassini-Huygens mission has been a remarkable success answering many outstanding questions as well as raising many new ones (Coustenis et al. 2009). It highlighted the complexity of Titan's atmospheric chemistry; however, the minimum flyby altitudes of 950 km limited the ability to explore the full set of chemical processes in the middle atmosphere where haze particles grow and evolve. The limited high-resolution spatial coverage limits our view of the range of detailed geological processes ongoing on this body. The exact chemical composition and structure of haze particles are still unknown, as well as the composition of surface materials and dissolved materials in the lakes. These would be one of the major targets in the next space mission to Titan, as several mission concepts have been planned (e.g., Reh et al. 2009; Coustenis et al. 2009; Stofan et al. 2013; Lorenz et al. 2017). Especially for an astrobiological context, identification and quantification of complex organic molecules in the vapor, condensed, and solid forms from ionosphere to the surface will be crucial. The chemical variations and their correlation to the geological processes are important. For example, to identify organic samples that appear to have been altered by liquid water might provide us an opportunity to understand a series of intermediate steps toward possible transition from abiotic to biotic processes. A search for homochirality and isotope signatures in organic matter is essential to a search of biological selection. A search for any biotic processes on Titan represents a test of life's cosmic ubiquity (Lunine 2009). Titan becomes an even more fascinating target for astrobiology than ever.

Acknowledgments HI is partially supported by the NASA Cassini Data Analysis Program. Dr. Mark Smith and an anonymous reviewer are acknowledged for valuable comments to improve the quality of this manuscript.

References

- Anderson CM, Samuelson RE (2011) Titan's aerosol and stratospheric ice opacities between 18 and 500 μm : vertical and spectral characteristics from Cassini CIRS. *Icarus* 212:762. <https://doi.org/10.1016/j.icarus.2011.01.024>
- Barnes JW, Soderblom JM, Brown RH et al (2009) VIMS spectral mapping observations of Titan during the Cassini prime mission. *Planet Space Sci* 57:1950–1962. <https://doi.org/10.1016/j.pss.2009.04.013>
- Béghin C, Canu P, Karkoschka E et al (2009) New insights on Titan's plasma-driven Schumann resonance inferred from Huygens and Cassini data. *Planet Space Sci* 57:1872. <https://doi.org/10.1016/j.pss.2009.04.006>
- Benner S, Ricardo A, Carrigan M (2004) Is there a common chemical model for life in the universe? *Curr Opin Chem Biol* 8:672–689
- Bezdard B (2009) Composition and chemistry of Titan's stratosphere. *Philos T R Soc A* 367:683–695. <https://doi.org/10.1098/rsta.2008.0186>
- Biemann K (2006) Astrochemistry: complex organic matter in Titan's aerosols? *Nature* 444:E6–E6
- Brown R, Lebreton J-P, Waite H (eds) (2009) Titan from Cassini-Huygens. Springer, New York
- Cable ML, Hörst SM, Hodyss R et al (2011) Titan Tholins: simulating Titan organic chemistry in the Cassini-Huygens era. *Chem Rev* 112:1882–1909. <https://doi.org/10.1021/cr200221x>
- Coustonis A, Atreya SK, Balint T et al (2009) TandEM: Titan and Enceladus mission. *Exp Astron* 23:893–946. <https://doi.org/10.1007/s10686-008-9103-z>
- Fulchignoni M, Ferri F, Angrilli F et al (2005) In situ measurements of the physical characteristics of Titan's environment. *Nature* 438:785–791. <https://doi.org/10.1038/nature04314>
- Hayes AG (2016) The lakes and seas of Titan. *Annu Rev Earth Planet Sci* 44:57–83. <https://doi.org/10.1146/annurev-earth-060115-012247>
- Horst SM (2017) Titan's atmosphere and climate. *J Geophys Res-Planet* 122:432–482. <https://doi.org/10.1002/2016JE005240>
- Iess L, Jacobson RA, Ducci M et al (2012) The tides of Titan. *Science* 337:457–459. <https://doi.org/10.1126/science.1219631>
- Imanaka H, Smith MA (2010) Formation of nitrogenated organic aerosols in the Titan upper atmosphere. *PNAS* 107:12423–12428. <https://doi.org/10.1073/pnas.0913353107>
- Imanaka H, Khare BN, Elsila J et al (2004) Laboratory experiments of Titan tholin formed in cold plasma at various pressures: implications for nitrogen-containing polycyclic aromatic compounds in Titan haze. *Icarus* 168:344–366. <https://doi.org/10.1016/j.icarus.2003.12.014>
- Imanaka H, Cruikshank DP, Khare BN, McKay CP (2012) Optical constants of Titan tholins at mid-infrared wavelengths (2.5–25 μm) and the possible chemical nature of Titan's haze particles. *Icarus* 218:247–261. <https://doi.org/10.1016/j.icarus.2011.11.018>
- Israel G, Szopa C, Raulin F et al (2005) Complex organic matter in Titan's atmospheric aerosols from in situ pyrolysis and analysis. *Nature* 438:796–799. <https://doi.org/10.1038/nature04349>
- Janssen MA, Le Gall A, Lopes RM et al (2016) Titan's surface at 2.18-cm wavelength imaged by the Cassini RADAR radiometer: results and interpretations through the first ten years of observation. *Icarus* 270:443. <https://doi.org/10.1016/j.icarus.2015.09.027>
- Koskinen T, Yelle R, Snowden D et al (2011) The mesosphere and lower thermosphere of Titan revealed by Cassini/UVIS stellar occultations. *Icarus* 216:507–534. <https://doi.org/10.1016/j.icarus.2011.09.022>

- Lavvas P, Yelle RV, Koskinen T et al (2013) Aerosol growth in Titan's ionosphere. *PNAS* 110:2729–2734. <https://doi.org/10.1073/pnas.1217059110>
- Lopes RMC, Kirk RL, Mitchell KL et al (2013) Cryovolcanism on Titan: new results from Cassini RADAR and VIMS. *J Geophys Res-Planet* 118:416–435. <https://doi.org/10.1002/jgre.20062>
- Lorenz RD, Wall S, Radebaugh J et al (2006) The sand seas of Titan: Cassini RADAR observations of longitudinal dunes. *Science* 312:724–727. <https://doi.org/10.1126/science.1123257>
- Lorenz RD, Mitchell KL, Kirk RL et al (2008) Titan's inventory of organic surface materials. *Geophys Res Lett* 35:L02206. <https://doi.org/10.1029/2007GL032118>
- Lorenz RD, Turtle EP, Barnes JW, et al (2017) Dragonfly: a Rotorcraft Lander Concept for scientific exploration at Titan. John Hopkins APL Technical Digest, PRE-PUBLICATION DRAFT, www.jhuapl.edu/techdigest
- Lunine JI (2009) Saturn's Titan: a strict test for life's cosmic ubiquity. *Proc Am Philos Soc* 153:403–418. <https://doi.org/10.2307/20721510>
- Lunine JI (2017) Ocean worlds exploration. *Acta Astronaut* 131:123–130. <https://doi.org/10.1016/j.actaastro.2016.11.017>
- Lunine JI, Atreya SK (2008) The methane cycle on Titan. *Nat Geosci* 1:159–164. <https://doi.org/10.1038/ngeo125>
- McKay CP (2016) Titan as the abode of life. *Lifestyles* 6:8–15. <https://doi.org/10.3390/life6010008>
- McKay CP, Smith H (2005) Possibilities for methanogenic life in liquid methane on the surface of Titan. *Icarus* 178:274–276. <https://doi.org/10.1016/j.icarus.2005.05.018>
- Mitchell JL, Lora JM (2016) The climate of Titan. *Annu Rev Earth Planet Sci* 44:353–380. <https://doi.org/10.1146/annurev-earth-060115-012428>
- Müller-Wodarg I, Griffith CA, Lellouch E, Cravens TE (eds) (2014) Titan: interior, surface, atmosphere, and space environment. Cambridge University Press, Cambridge
- Neish CD, Barnes JW, Sotin C et al (2015) Spectral properties of Titan's impact craters imply chemical weathering of its surface. *Geophys Res Lett* 42:3746–3754. <https://doi.org/10.1002/2015GL063824>
- Niemann H, Atreya S, Bauer S et al (2005) The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* 438:779–784. <https://doi.org/10.1038/nature04122>
- Porco C, Baker E, Barbara J et al (2005) Imaging of Titan from the Cassini spacecraft. *Nature* 434:159–168. <https://doi.org/10.1038/nature03436>
- Raulin F, Brassé C, Poch O, Coll P (2012) Prebiotic-like chemistry on Titan. *Chem Soc Rev* 41:5380. <https://doi.org/10.1039/c2cs35014a>
- Reh K, Magner T, Matson D et al (2009) Titan Saturn system mission study 2008: final report. Jet Propulsion Laboratory, Pasadena
- Sagan C, Khare B, Lewis J et al (1984) Organic matter in the Saturn system. In: Gehrels T, Matthews MS (eds) *Saturn*. University of Arizona Press, Tucson, pp 788–807
- Soderblom LA, Kirk RL, Lunine JI et al (2007) Correlations between Cassini VIMS spectra and RADAR SAR images: implications for Titan's surface composition and the character of the Huygens probe landing site. *Planet Space Sci* 55:2025–2036. <https://doi.org/10.1016/j.pss.2007.04.014>
- Stevenson J, Lunine J, Clancy P (2015) Membrane alternatives in worlds without oxygen: creation of an azotosome. *Sci Adv* 1:e1400067. <https://doi.org/10.1126/sciadv.1400067>
- Stofan ER, Elachi C, Lunine JI et al (2007) The lakes of Titan. *Nature* 445:61–64. <https://doi.org/10.1038/nature05438>
- Stofan E, Lorenz R, Lunine J, et al (2013) TiME-the Titan Mare Explorer. Aerospace Conference, 2013 IEEE 1–10. doi: <https://doi.org/10.1109/AERO.2013.6497165>
- Strobel DF (2010) Molecular hydrogen in Titan's atmosphere: implications of the measured tropospheric and mesospheric mole fractions. *Icarus* 208:878–886. <https://doi.org/10.1016/j.icarus.2010.03.003>

- Tobie G, Lunine J, Sotin C (2006) Episodic outgassing as the origin of atmospheric methane on Titan. *Nature* 440:61–64. <https://doi.org/10.1038/nature04497>
- Tomasko M, Archinal B, Becker T et al (2005) Rain, winds and haze during the Huygens probe's descent to Titan's surface. *Nature* 438:765–778. <https://doi.org/10.1038/nature04126>
- Waite JH, Young DT, Cravens TE et al (2007) The process of tholin formation in Titan's upper atmosphere. *Science* 316:870–875. <https://doi.org/10.1126/science.1139727>

Chapter 27

Panspermia Hypothesis: History of a Hypothesis and a Review of the Past, Present, and Future Planned Missions to Test This Hypothesis



Yuko Kawaguchi

Abstract Speculations about the origins of life on Earth have existed since the dawn of civilization. The Greek philosopher Anaxagoras (500–428 BCE) asserted that the seeds of life are present everywhere in the universe (Nicholson, *Trends Microbiol* 17:243–250, 2009). He coined the term panspermia to describe the concept as life traveling between planets as seed. The other Greek philosophers, Anaximander (588–524 BCE) and Thales (624–548 BCE), mentioned philosophical point of panspermia theory. Many famous nineteenth-century scientists also wrote about this theory. Among others, Svante Arrhenius posited that microscopic spores are transferred through interplanetary space by means of radiation pressure from the sun, in 1903. In the modern formulation, there are three stages envisioned in this hypothesis: escape (from a planet), transit (through interplanetary space), and landing (on a recipient planet). Each stage has since been investigated, lending some credence to the hypothesis. For example, the possibility of microbial spores escaping a planet has been supported by the capture of radioresistant microbes from high altitudes on Earth. From the space experiments conducted in Earth orbiters and on the International Space Station (ISS), microbes have been found to survive at low Earth orbits (LEO) under some protection from intense solar UV radiation, which could well be available for spores embedded within meteorites. Heating up in the atmosphere due to friction is the main problem during reentry to the planet with atmosphere. However, because the time spent under intense friction is generally in the order of only a few tens of seconds, the amount of heat generated may not be sufficient to kill all the spores, especially if hitching a ride within meteorites. The panspermia hypothesis has been modified and revived since its original proposal and has given a new perspective to the explorations on Mars or the icy moons of Jupiter and Saturn. The hypothesis, its modifications, and past and ongoing research are reviewed in this chapter.

Y. Kawaguchi (✉)

Planetary Exploration Research Center (PERC), Chiba Institute of Technology (CIT),
Narashino, Chiba, Japan

e-mail: kawaguchi@perc.it-chiba.ac.jp

Keywords Panspermia · Microbes · Space experiment · Low Earth orbit

27.1 Introduction

Speculations about the origins of life on Earth have existed since the dawn of civilization. The Greek philosopher Anaxagoras (500–428 BCE) asserted that the seeds of life are present everywhere in the universe (Nicholson 2009). He coined the term panspermia to describe the concept as life traveling between planets as seed. The other Greek philosophers, Anaximander (588–524 BCE) and Thales (624–548 BCE), also mentioned philosophical point of panspermia theory. Many famous nineteenth-century scientists also wrote about this theory. Among others, the Swedish chemist and Nobel laureate Svante Arrhenius published *Worlds in the Making: The Evolution of the Universe* in 1903, suggesting that microscopic spores were transported through interplanetary space by means of radiation pressure from the sun (Arrhenius 1903). The hypothesis is called panspermia (“pan” and “sperma” mean “all” and “seed,” respectively, in Greek). It argues that terrestrial life was seeded by extraterrestrial life forms that traveled to Earth by radiation pressure (Weber and Greenberg 1985) or within meteorites (Melosh 1988). This seeding of Earth could have been the result of directed panspermia, which is defined as the deliberate seeding of one planet by the civilizations in another, for example, the seeding of Earth from outer space by alien civilizations (Crick and Orgel 1973; Hoyle and Wickramasinghe 1979) or the seeding of extraterrestrial bodies from Earth (Mautner and Matloff 1979). The problem of UV radiation in interplanetary space, which can be lethal to microbial survival, is thought to be nullified or at least minimized, when microbial spores travel within meteorites – a process called *litho-panspermia*. This is considered the most likely scenario as the interplanetary exchange of rocks (meteorites) is a natural and ongoing phenomenon, and meteorites that do not burn up in the atmosphere shatter upon impact, spewing dust and broken-off chunks into the host planet’s atmosphere and soil, including any microbial spores present within.

There are three stages to the process of panspermia: (I) escaping from donor planet, (II) transport through space from donor planet to recipient planet, and (III) landing on recipient planet (Fig. 27.1). Several different hazards for microbial survival at each step have been investigated (reviewed by Horneck et al. 2002; Nicholson 2009). We discuss each stage for the possibility and viability of interplanetary transfer of microbes.

27.2 Escape from Donor Planet

Capture experiments of microbes at high altitudes have been performed on Earth using aircrafts, observation balloons, and meteorological rockets (reviewed in Griffin 2004; Kellogg and Griffin 2006; Yang et al. 2009a; Smith 2013; Kawaguchi

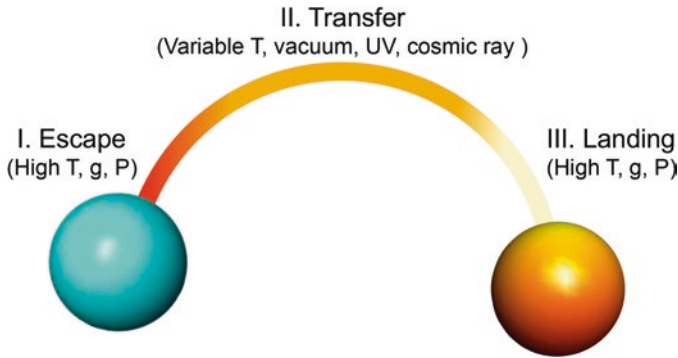


Fig. 27.1 Scenario of panspermia. (I) Escape from a donor planet at high temperature (T), gravity (g), and atmospheric pressure (P). (II) An object traveling through interplanetary space is exposed to variable T, vacuum, UV, and cosmic rays. (III) The object is captured by a recipient planet, enters the atmosphere, and falls to impact the surface with high T, g, and P. (Modified from Nicholson 2009)

et al. 2016). Hypotheses for the transportation of microbes to high altitudes typically involve the action of bioaerosols. For example, thunderstorms (Dehel et al. 2008), volcanic eruptions (Griffin 2004; Van Eaton et al. 2013), and giant meteorite impacts (Kring 2000; Mileikowsky et al. 2000; Gladman et al. 2005; Worth et al. 2013) are considered to be responsible for the transportation of bioaerosols across the troposphere. Human activity (e.g., airplanes, balloons, rockets, and spacecrafts) are also considered possible means of transport of bioaerosols from the ground to the upper atmosphere (Bucker and Horneck 1968; Griffin 2004; Smith 2013). It is also hypothesized that the electric forces of transient luminous events such as sprites, gigantic jets, and blue jets in the high atmosphere accelerate bioaerosols that as a result are transported over the stratosphere or mesosphere (Kawaguchi et al. 2016).

Microbe capture experiments at high altitudes have been reported. Yang et al. (2008) reported that an air-sampler collected dust on membrane filters from the low stratosphere and high troposphere and that radiation-resistant *Deinococcus* sp. was cultured when inoculated with the membrane filters (Yang et al. 2009b, 2010). Capture experiments conducted at an altitude of 20 km using NASA's aircraft flying over the American continent (Griffin 2004, 2008; Smith et al. 2010) exposed sterilized impactor plates outside of aircraft during the flight. The impactor plates were placed on a R2A medium, and spore-forming *Bacillus* sp. and nonspore-forming bacteria were identified. Balloons and rockets are able to reach altitudes higher than that of aircraft, such as the capture experiment at altitudes of 48–77 km using a rocket (Imshenetsky et al. 1978). Mainly radiation-resistant bacteria and fungi have been isolated in previous dust sampling experiments (Kawaguchi et al. 2016). However, the microbes discovered also depend on the culture methods, and culture conditions are typically different in different experiments. It is also suggested that

cultivable microbes are less than 0.1~0.001% of the possible microbes present (Amann et al. 1995).

Some isolated microbes were reported to show the tendency to form cell aggregates within particles (e.g., rock fragments) (Lighthart 1997; Harris et al. 2002; Wainwright et al. 2004; Yang et al. 2008). Intense UV radiation is lethal to naked microbial cells. Microbes inside aggregates or within rock could be protected against intense UV at high altitudes and are likely to survive longer during transportation. Based on capture experiments, it appears that the microbial density (in colony-forming units, CFU) depends on the altitude from which the microbes were captured (Yang et al. 2009a; Kawaguchi et al. 2016) (Fig. 27.2). The figure indicates that CFU decreases with increasing altitude. However, no CFU data obtained from capture experiments of microbes over the stratosphere have been reported.

To investigate the boundary of the Earth's biosphere and the possibility that microbes escape from Earth to space, a Japanese astrobiology experiment named Tanpopo mission is in progress. Terrestrial microparticles are captured outside of the International Space Station (ISS) (Yamagishi 2007; Kawaguchi et al. 2016). Blocks of silica aerogels are exposed, which capture orbiting microparticles. Impact tracks and particles are stained with DNA-specific fluorescence dye to identify terrestrial microbes in the aerogels (Kawaguchi et al. 2014). The analysis has been performed since 2016. If the microbes are found in the ISS orbit, the results will

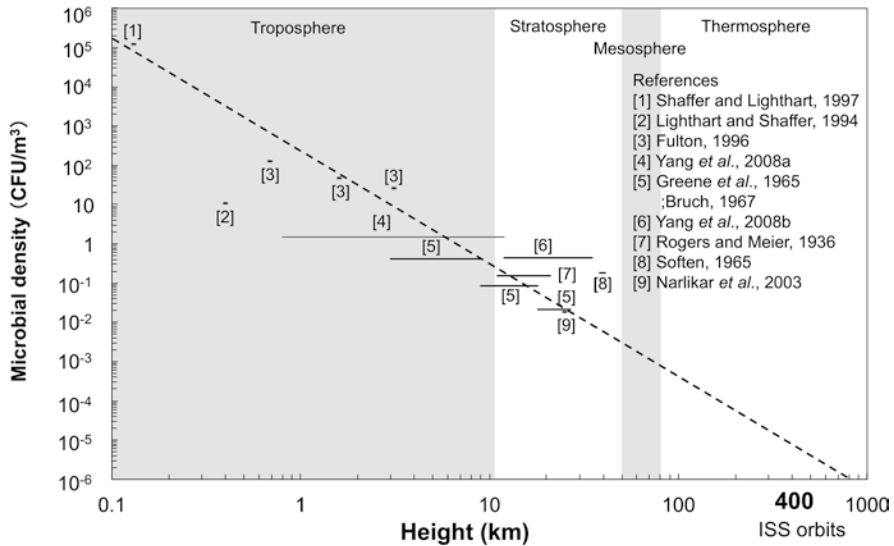


Fig. 27.2 The microbial density (cfu) depending on the height created by the results of capture experiments of microbes at high atmosphere. Horizontal lines or filled circles were drawn based on the data at ambient temperature and pressure from each reference. Because some microbial sampling studies were carried out in certain altitude ranges, the horizontal lines show the respective sampling altitude ranges. The inserted dashed line indicates the estimated microbial density versus height dependence. (Modified from Kawaguchi et al. 2016 with permission from Mary Ann Liebert, Inc.)

expand the limit of the terrestrial biosphere and further bolster the panspermia hypothesis.

27.3 Microbial Survival in Space

Microbial survival in space could not be tested until the second half of the twentieth century. Since the 1960s, however, exposure experiments of microbes have been conducted to test their survival at low Earth orbit (LEO) using Earth orbiters, the Russian manned spacecraft MIR, space shuttle, and ISS (Table. 27.1, Fig. 27.3). Previous and current microbe-exposure experiments and their facilities have been summarized in Taylor (1974), Horneck et al. (2010), and Cottin et al. (2017).

Table 27.1 Summary of exposure experiment of microbes in space

Year	Mission	Exposed microbes	Result	Ref.
1960	Rockets; altitude of 150 km	Bacteriophage T1	Exposed at 150 km for 3 min: killed	Hotchin et al. (1968)
		<i>B. subtilis</i> spores		
		<i>Penicillium</i> spores		
1983	Spacelab; altitude of 240 km	<i>B. subtilis</i> spores	Single layer of spores: killed by UV irradiation	Horneck et al. (1984)
1984–1990	LDEF; altitude ~500 km	<i>B. subtilis</i> spores	Multilayer of spores: survived	Horneck et al. (1994)
1999	FOTON, Biopan	<i>B. subtilis</i> spores	Spores with clay and glucose as “mixed layers”: survived for 21 days	Horneck et al. (2001)
2008–2009	EXPOSE-E, ISS	Lichen <i>Xanthoria elegans</i>	Rock-colonizing lichen: survived for 1.5 years	Onofri et al. (2012)
		<i>Rhizocarpon geographicum</i>		
2014–2016	EXPOSE-R2, ISSISS	<i>D. geothermalis</i>	Biofilm survived for 16 months at ISS	Baquéq et al. (2013)
		<i>Chroococcidiopsis</i>		Frösler et al. (2017)
				Bill et al. (2017)
2015–2018	Tanpopo, ISS	<i>D. radiodurans</i> R1		Kawaguchi et al. (2013)
		<i>D. aerius</i>		
		<i>D. aetherius</i>		
		DNA repair gene-deficient mutants in <i>D. radiodurans</i> R1		

Modified from Horneck et al. (2010)

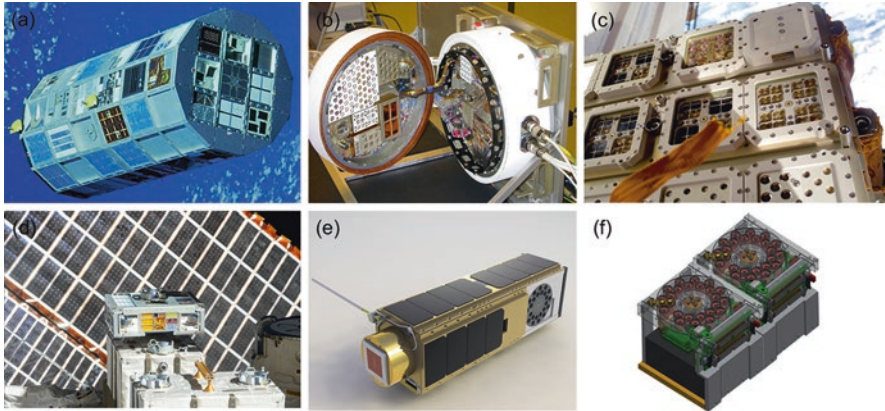


Fig. 27.3 Some of the past and current facilities and devices used for microbial exposure experiments. (a) LDEF, (b) FOTON with embedded samples in its heat shield for STONE experiment, (c) EXPOSE-E at outside of European module of ISS, (d) Tanpopo outside of Kibo-ISS, (e) O/OREOS and (f) OREOcube. (Credits: NASA, ESA, ESA/NASA, JAXA/NASA, NASA Ames Research Center)

The first experiment in LEO to test the survival of microbes was performed to expose bacteriophage T1, *Bacillus* spores, and *Penicillium* spores for 3 min at an altitude of 150 km using the Gemini satellite (Hotchin et al. 1968). The exposed microbes were found to have been inactivated. In the later experiments, dried spores of *Bacillus subtilis* were placed on a slide beneath the aluminum dome either without cover or covered by quartz filters or aluminum. *Bacillus subtilis* cells exposed to UV in a single layer without cover were killed by solar UV. However a multilayer of bacillus spores survived for 6 years, in the longest exposure experiment to date, at altitude of ~500 km in the Long Duration Exposure Facility, LDEF (Horneck et al. 1994) (Fig. 27.3). *Bacillus* spores with clay and glucose in “mixed layers” survived for 21 days using the BIOPAN facility of the European Space Agency (ESA) onboard a Russian FOTON satellite (Horneck et al. 2001). Rock-colonizing lichen *Xanthoria elegans* and *Rhizocarpon geographicum* showed high survival fractions for 1.5 years at the ISS onboard the European facility in space mission EXPOSE-E (Onofri et al. 2012). During the exposure experiment, rock-colonizing cells were exposed to the full space environment (vacuum from 10^{-7} to 10^{-4} Pa, fluctuations of temperature between -21 °C and $+59.6$ °C, cosmic ionizing radiation up to 190 mGy, and solar UV up to 6.34×10^8 J/m²) (Rabbow et al. 2012). These results suggest that extremophiles can survive inside a protection that shields them from the intense solar UV radiation. The rocky panspermia is named lithopanspermia where “litho” stands for rock or stone (e.g., Melosh 1988; Horneck et al. 2002; Paniet et al. 2015). Such a transfer of microbes inside of meteoroids is expected to have been occurred mainly during the Late Heavy Bombardment.

The time needed for meteoroids to be transferred between Mars and Earth is estimated a few months to years depending on the orbit (Melosh 1988; Mileikowsky

et al. 2000). Mars is believed to have had life-supporting environmental conditions (e.g., water) in the past (Squyres et al. 2004). Fossils of nano size bacteria-like structures were observed at the surface of the Martian meteorite ALH84001 (McKay et al. 1996). However, the credibility of interpretation of the structure as fossilized organism is not established, yet. The best remaining evidence from ALH84001 is the magnetite particles in ALH84001, which are similar to terrestrial magnetite particles known as magnetosomes (Weiss et al. 2004). There has been no evidence of lithopanspermia from meteorites. However, simulation studies showed that giant impacts can transport a number of ejectiles from Mars to Earth or Earth to Mars and beyond, reaching the moons of Jupiter or Saturn (Mileikowsky et al. 2000; Worth et al. 2013).

The other possible protection from UV radiation is by means of biofilms or cell aggregates. A European group has investigated the survival of *Deinococcus geothermalis* and *Chroococcidiopsis* under the protection of biofilms in the missions Biofilm Organisms Surfing Space (BOSS) and EXPOSE-R2 (Baque et al. 2013; Frösler et al. 2017). Survival of *D. radiodurans* R1 and *D. aetherius* cell aggregates under space environment has been tested in the Japanese space mission Tanpopo on the Exposure Facility of the Japanese Experimental Module of ISS (Kawaguchi et al. 2013). The results suggest that the submillimeter level cell aggregates can be transferred from Earth to other planets. The hypothesis was named the massapanspermia hypothesis (Kawaguchi et al. 2013).

27.4 Survival of the Landing Process

The landing process is well studied under the lithopanspermia theory. When a recipient planet with atmosphere captures a rock, the rock reaches very high temperatures during landing. However, since the fall through the atmosphere takes only a few seconds, the outer layers of a rock can protect inner parts against heat (Horneck et al. 2001). It was found that only 3 mm of the surface of the Martian meteorite ALH84001 melted and formed a fusion crust (Weiss et al. 2000). Actually, for non-metallic meteorites, the heat of reentry is taken away by melt droplets from the surface faster than the heat pulse can travel inward, leaving the interior cool. The inner parts of the meteorite had not been heated to more than 40 °C during ejection from Mars and landing on Earth through the atmosphere (Weiss et al. 2000).

After the space trip and landing on Mars, microorganisms would need to overcome and maintain their life under extreme conditions. Exposure experiments of microbes are in progress on ISS investigating microbial survival under Mars like conditions using cultivation-dependent and cultivation-independent techniques by the ESA team. The NASA cubeSat Organism/Organic Exposure to Orbital Stresses (O/OREOS) allows the collection of data on microbial survival and metabolic activity and the investigation of microbes and biomarkers in situ under space conditions when orbiting Earth (Nicholson et al. 2011). OREOcube is also a cubeSat-based space exposure platform with in situ spectroscopy capabilities, which allows

the investigation of the origin and evolution of organic molecules in space and planetary environments, developed by ESA (Elsaesser et al. 2014). These next-generation platforms of exposure facilities enable mid-infrared diagnostics for more sophisticated experiments.

27.5 Conclusion

Development of space technology allows us to investigate the panspermia hypothesis. However, we are still in the process of determining if panspermia occurred in interplanetary space, in the past or at present. Next-generation explorations and current space missions will be instrumental in investigating the panspermia theory. If these future explorations find panspermia plausible, the results will profoundly affect the interpretation of the future space life search projects.

References

- Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59:143–169
- Arrhenius S (1903) Die Verbreitung des Lebens im Welten-raum. *Unschau* 7:481–485
- Baque´ M, Scalzi G, Rabbow E, Rettberg P, Billi D (2013) Biofilm and planktonic lifestyles differently support the resistance of the desert cyanobacterium *Chroococcidiopsis* under space and martian simulations. *Orig Life Evol Biosph* 43:377–389
- Baqueq M, Scalzi G, Rabbow E et al (2013) Biofilm and planktonic lifestyles differently support the resistance of the desert cyanobacterium *Chroococcidiopsis* under space and Martian simulations. *Orig Life Evol Biosph* 43:377–389
- Bill D, Verseux C, Rabbow E et al (2017) Endurance of desert-cyanobacteria biofilms to space and simulated Mars conditions during the EXPOSE-R2 space mission. In: Abstracts of European Astrobiology Network Association 2017, Aarhus University, Denmark, 14–18 August 2017
- Bucker H, Horneck G (1968) Discussion of a possible contamination of space with terrestrial life. *Life Sci Space Res* 7:21–27
- Cottin H, Kotler JM, Billi D et al (2017) Space as a tool for astrobiology: review and recommendations for experimentations in Earth orbit and beyond. *Space Sci Rev* 209(1–4):83–181
- Crick FHC, Orgel LE (1973) Directed panspermia. *Icarus* 19:341–346
- Dehel T, Lorge F, Dickinson M (2008) Uplift of microorganisms by electric fields above thunderstorms. *J Electrostat* 66:463–466
- Elsaesser A, Quinn RC, Ehrenfreund P et al (2014) Organics Exposure in Orbit (OREOCube): a next-generation space exposure platform. *Langmuir* 30:13217–13227
- Frösler J, Panitz C, Wingender J et al (2017) Survival of *Deinococcus geothermalis* in biofilms under desiccation and simulated space and Martian conditions. *Astrobiology* 17:431–447
- Gladman B, Dones L, Levison HF et al (2005) Impact seeding and reseedling in the inner solar system. *Astrobiology* 5:483–496
- Griffin DW (2004) Terrestrial microorganisms at an altitude of 20,000m in Earth’s atmosphere. *Aerobiologia* 20:135–140
- Griffin DW (2008) Non-spore-forming eubacteria isolated at an altitude of 20,000m in Earth’s atmosphere: extended incubation periods needed for culture-based assays. *Aerobiologia* 24:19–25

- Harris MJ, Wickramasinghe NC, Lloyd D et al (2002) The detection of living cells in stratospheric samples. *Proc SPIE* 4495:192–198
- Horneck G, Bücker H, Dose K et al (1984) Microorganisms and biomolecules in space environment, experiment ES029 on Spacelab 1. *Adv Space Res* 4:19–27
- Horneck G, Bucker H, Reitz G (1994) Long-term survival of bacterial spores in space. *Adv Space Res* 14:41–45
- Horneck G, Rettberg P, Reitz G et al (2001) Protection of bacterial spores in space, a contribution to the discussion on panspermia. *Orig Life Evol Biosph* 31:527–547
- Horneck G, Mileikowsky C, Melosh HJ et al (2002) Viable transfer of microorganisms in the solar system and beyond. In: Horneck G, Baumstark-Khan C (eds) *Astrobiology: the quest for the conditions of life*. Springer, Berlin, pp 57–76
- Horneck G, Klaus DM, Mancinelli RL (2010) Space microbiology. *Microbiol Mol Biol Rev* 74:121–156
- Hotchin J, Lorenz P, Hemenway C (1968) The survival of terrestrial microorganisms in space at orbital altitudes during Gemini satellite experiments. *Life Sci Space Res* 6:108–114
- Hoyle F, Wickramasinghe NC (1979) *Diseases from space*. Dent, London
- Imshenetsky A, Lysenko S, Kazakov G (1978) Upper boundary of the biosphere. *Appl Environ Microbiol* 35:1–5
- Kawaguchi Y, Yang Y, Kawashiri N et al (2013) The possible interplanetary transfer of microbes: assessing the viability of *Deinococcus* spp. under the ISS environmental conditions for performing exposure experiments of microbes in the Tanpopo mission. *Orig Life Evol Biosph* 43:411–428
- Kawaguchi Y, Sugino T, Tabata M et al (2014) Fluorescence imaging of microbe-containing particles shot from a two-stage light-gas gun into an aerogel. *Orig Life Evol Biosph* 44:43–60
- Kawaguchi Y, Yokobori S, Hashimoto H, Yano H, Tabata M et al (2016) Investigation of the interplanetary transfer of microbes in the Tanpopo mission at the exposed facility of the international space station. *Astrobiology* 16:363–376
- Kellogg CA, Griffin DW (2006) Aerobiology and the global transport of desert dust. *Trends Ecol Evol* 21:638–644
- Kring DA (2000) Impact events and their effect on the origin, evolution, and distribution of life. *GSA Today* 10:1–7
- Lighthart B (1997) The ecology of bacteria in the alfresco atmosphere. *FEMS Microbiol Ecol* 23:263–274
- Mautner M, Matloff G (1979) Directed panspermia: a technical evaluation of seeding nearby solar systems. *J Br Interplanet Soc* 32:419–422
- McKay DS, Gibson EK Jr, Thomas-Keprta KL et al (1996) Search for past life on Mars: possible relic biogenic activity in Martian meteorite ALH 84001. *Science* 273:924–930
- Melosh HJ (1988) The rocky road to panspermia. *Nature* 332:687–688
- Mileikowsky C, Cucinotta FA, Wilson JW et al (2000) Natural transfer of viable microbes in space—1. From Mars to Earth and Earth to Mars. *Icarus* 145:391–427
- Nicholson WL (2009) Ancient micronauts: interplanetary transport of microbes by cosmic impacts. *Trends Microbiol* 17:243–250
- Nicholson WL, Ricco AJ, Agasid E et al (2011) The O/OREOS Mission: first science data from the Space Environment Survivability of Living Organisms (SESLO) payload. *Astrobiology* 11:951–958
- Onofri S, de la Torre R, de Vera J-P et al (2012) Survival of rock-colonizing organisms after 1.5 years in outer space. *Astrobiology* 12:508–516
- Panitz C, Horneck G, Rabbow E et al (2015) The SPORES experiment of the EXPOSE-R mission: *Bacillus subtilis* spores in artificial meteorites. *Int J Astrobiol* 14:105–114
- Panitz C, Frösler J, Walingender J et al (2017) The BOSS experiment of the EXPOSE-R2 mission: biofilms versus planktonic cells. In: Abstracts of European Astrobiology Network Association 2017, Aarhus University, Denmark, 14–18 August 2017

- Rabbow E, Rettberg P, Barczyk S, Bohmeier M, Parpart A, Panitz C, Horneck G, von Heise-Rotenburg R, Hoppenbrouwers T, Willnecker R, Baglioni P, Demets R, Dettmann J, Reitz G (2012) EXPOSE-E: an ESA astrobiology mission 1.5 years in space. *Astrobiology* 12:374–386
- Smith DJ (2013) Microbes in upper atmosphere and unique opportunity for astrobiology research. *Astrobiology* 13:981–990
- Smith DJ, Griffin DW, Schuerger AC (2010) Stratospheric microbiology at 20km over the Pacific ocean. *Aerobiologia* 26:35–46
- Squyres SW, Grotzinger JP, Arvidson RE et al (2004) In situ evidence for an ancient aqueous environment at Meridiani Planum. *Mar Sci* 306:1709–1714
- Taylor GR (1974) Space microbiology. *Annu Rev Microbiol* 28:121–137
- Van Eaton AR, Harper MA, Wilson CJN (2013) High-flying diatoms: widespread dispersal of microorganisms in an explosive volcanic eruption. *Geology* 41:1187–1190
- Wainwright M, Wickramasinghe NC, Narlikar JV et al (2004) Confirmation of the presence of viable but noncultureable bacteria in the stratosphere. *Int J Astrobiol* 3:13–15
- Weber P, Greenberg JM (1985) Can spores survive in interstellar space? *Nature* 316:403–407
- Weiss BP, Kirschvink JL, Baudenbacher FJ et al (2000) A low temperature transfer of ALH84001 from Mars to Earth. *Science* 290:791–795
- Weiss BP, Kim SS, Kirschvink JL, Kopp RE, Sankaran M, Kobayashi A, Komeili A (2004) Magnetic tests for magnetosome chains in Martian meteorite ALH84001. *Proc Natl Acad Sci U S A* 101(22):8281–8284. Epub 2004 May 20
- Worth RJ, Sigurdsson S, House CH (2013) Seeding life on the moons of the outer planets via lithopanspermia. *Astrobiology* 13:1155–1165
- Yamagishi A (2007) Tanpopo: astrobiology exposure and micrometeoroid capture experiments on the EUSO. *Biol Sci Space* 21:67–75
- Yang Y, Itahashi S, Yokobori S et al (2008) UV-resistant bacteria isolated from upper troposphere and lower stratosphere. *Biol Sci Space* 22:18–25
- Yang Y, Yokobori S, Yamagishi A (2009a) Assessing panspermia hypothesis by microorganisms collected from the high altitude atmosphere. *Biol Sci Space* 23:151–163
- Yang Y, Itoh T, Yokobori S et al (2009b) *Deinococcus aerius* sp. nov., isolated from the high atmosphere. *Int J Syst Evol Microbiol* 59:1862–1866
- Yang Y, Itoh T, Yokobori S et al (2010) *Deinococcus aetherius* sp. nov., isolated from the stratosphere. *Int J Syst Evol Microbiol* 60:776–779

Chapter 28

Extrasolar Planetary Systems



Motohide Tamura

Abstract The observational exploration of extrasolar planets or exoplanets is one of the hottest topics in modern astronomy. Over the last two decades, thousands of exoplanets have been discovered. Though most of these are within our Milky Way galaxy and relatively close to us, some may be beyond our Galaxy. Some are small planets of even sub-Earth sizes. Some have two “Suns.” Some are Earth-like rocky planets in the habitable zone (Kasting JF, Whitmire DP, Reynolds RT, *Icarus* 101:108–128, 1993; Kopparapu RK, *ApJ* 767:article id. 131, 2013), and some are temperate planets around red dwarfs whose environment is alien to us. We now know that most stars, not only Sun-like stars but also low-mass red dwarfs, host a system of one or more planets. The most spectacular aspect beyond our imagination is the diversity of exoplanets whose physical characters are very different from those of our solar system planets. Various exoplanet detection and characterization methods, both classical and new ones, have been applied in indirect and direct ways. Not only the fundamental planetary parameters such as mass, radius, and orbits but also planetary atmospheric information can now be obtained for the planets down to almost Earth size. This chapter summarizes the current knowledge on exoplanets and their detection method including some future plans as well as short introduction of the presently most interesting planets for astrobiology toward the detection and characterization of life-harboring exoplanets.

Keywords 51 Peg · 51 Eri · Beta Pic · Direct imaging · Doppler method · Eccentric planets · ELT · Exoplanet · Extrasolar planet · GJ 504 · GMT · HabEx · Habitable zone · High-contrast · HD 209458 · HR 8799 · Hot Jupiter · IFU · IRD · Kepler mission · LUVOIR · PDS 70 · PLATO · Proxima Centauri · Red dwarf · Subaru telescope · Super-Earth · TESS · TMT · Transit method · TRAPPIST-1 · TTV · WFIRST · Wide-orbit planets

M. Tamura (✉)

Department of Astronomy, Graduate School of Science, The University of Tokyo, Tokyo, Japan

Astrobiology Center, National Institutes of Natural Sciences, Tokyo, Japan
e-mail: motohide.tamura@nao.ac.jp

28.1 Discovery of Exoplanets and Their Varieties

Since the first discovery of planets around a normal star outside our solar system (Mayor and Queloz 1995) and a neutron star (Wolszczan and Frail 1992), nearly 4000 confirmed planets have been reported (see <http://exoplanet.eu/> for a catalog). These are called extrasolar planets or exoplanets. Not only this large number of discovered exoplanets but also the existence of various types of exoplanets has stimulated the recent development of the astrobiology field. Small planets with water such as Earth-like ones and super-Earths are the most promising sites for bearing life. Since the planets are formed as by-products of star formation processes, the observations of exoplanets and their formation site, protoplanetary disks, are also important for understanding the formation of a planetary system around a central star or central multiple stars.

Figure 28.1 summarizes the distribution of exoplanets discovered by various techniques as of February 2018 as well as three representative planets in our solar system. Our solar system consists of the Sun and eight planets. Mercury, Venus, Earth, and Mars are situated at 0.4–1.5 au, where 1 au is the mean distance between the Sun and Earth ($\sim 1.5 \times 10^{11}$ m). These are the lowest-mass planets (0.06 – $1.0 M_{\text{Earth}}$) and are mainly composed of rocks, thus called rocky planets or Earth-like planets. Jupiter and Saturn are situated in the middle of the solar system (5.2 and 9.6 au) and are the most massive planets (320 and 95 M_{Earth}). They are mainly composed of gas and are thus called gas giant planets, or Jovian planets. Uranus and Neptune are the furthest planets (19 and 30 au) and are of medium mass (15 and 17 M_{Earth}). They are mainly composed of ice and are thus called icy giant planets, or Neptunian planets.

New types of planets discovered for the first time in exoplanetary systems are as follows: (1) “hot Jupiters” whose masses comparable to Jupiter and whose tempera-

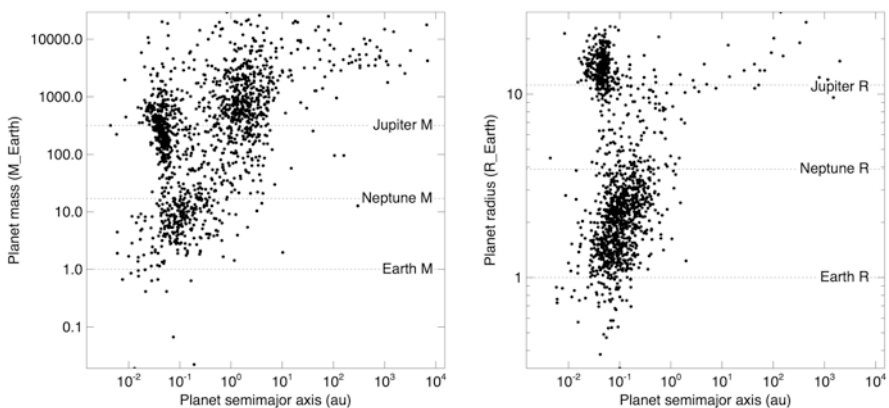


Fig. 28.1 Distribution of exoplanets discovered by various detection techniques as of February 2018. Left: Planet mass vs. planet semimajor axis for the confirmed planets with mass data regardless of their errors. Right: Planet radius vs. planet semimajor axis for the confirmed planets with radius data regardless of their errors. The planet mass/radius values for Earth, Neptune, and Jupiter with their semimajor axes of 1, 30, and 5 au, respectively, are also shown as vertical lines

tures are >1000 K because they are very close to their host stars ($a < 0.1$ au), (2) “super-Earths” whose sizes or masses are between those of Earth and Neptune, (3) “eccentric planets” whose orbital eccentricities are much larger than those of the solar system planets, and (4) “wide-orbit giant planets” whose orbit is beyond that of Neptune. Although many theories have been proposed to explain the formation of these exoplanets, no “universal” theory has been established yet.

The large number of the discovered exoplanets has also enabled us to discuss their demographics: almost all stars have a planet or planets (Cassan et al. 2012). Among them, the small planets such as Neptunian, super-Earth, and Earth-sized planets are dominant. However, these results are mainly based on the statistics of inner ($a < 0.4$ au or $P < 80$ days) planets discovered by the Kepler mission (see Sect. 28.2.3). Before going to more details of the exoplanetary systems, we first summarize the astronomical observation techniques for detecting and characterizing exoplanets. See several textbooks and a handbook for more details (e.g., Haswell 2010; Seager 2011; Perryman 2014; Tamura 2015).

28.2 Exoplanet Detection Methods

28.2.1 Doppler Method

The radial velocity (RV) or Doppler method measures small and periodic Doppler shifts of the absorption lines or bands of the host star due to the planetary revolution (see Fig. 28.2). The wavelength shifts are measured with a precise RV instrument or high-dispersion spectrometer. Its velocity amplitude K is described as:

$$K [\text{cm} / \text{s}] \sim 10 \left(M_{\text{planet}} [M_{\text{Earth}}] \times \sin i \right) / \left(a [\text{au}] \times M_{\text{star}} [M_{\text{Sun}}] \right)^{1/2}$$

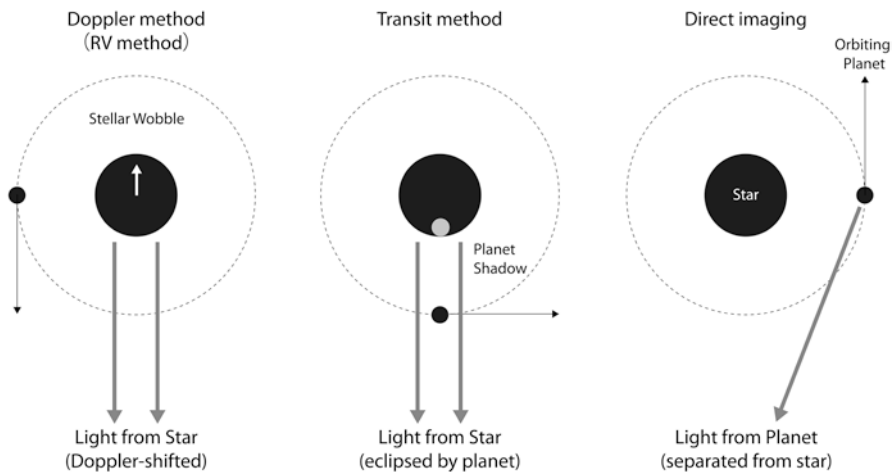


Fig. 28.2 Principles of the three exoplanet detection methods: the Doppler, transit, and direct imaging methods. Only primary transit case is shown in the transit figure. See text for more details

where M_{planet} is the planet mass, i is the orbital plane inclination, M_{star} is the stellar mass, M_{Earth} is the Earth mass, a is the semimajor axis, and M_{Sun} is the mass of the Sun. The Doppler method has a bias to detect massive, close-in planets. Note that the derived mass ($M_{\text{planet}} \sin i$) depends on the orbital inclination and therefore is the minimum mass. If the planetary system is seen exactly face-on, this method cannot detect the stellar wobble along the line of sight. However, this defect has a minor effect for the statistical discussion using their masses because the inclination-averaged mass is $\pi/4$ of the true mass.

One of the most critical techniques of this method is the wavelength calibration. Two types of the calibration methods have been used so far: the gas-cell (most frequently I^2 cell) technique and the lamp (most frequently ThAr lamp) technique. Recently, laser frequency comb technique is introduced to achieve the RV measurement accuracy of better than 1 m/s. Note that the RV amplitudes K of the Sun due to the orbital motions of Jupiter and Earth are approximately 12 m/s and 10 cm/s, respectively. Therefore, an Earth twin's RV amplitude around Sun-like stars is the order of 10 cm/s, which is challenged by stellar jitter and instrumental noises. In contrast, this is mitigated for Earth-sized planets on temperate orbits around M stars, red dwarfs, due to the small stellar mass and the small orbital semimajor axis.

The first successful and convincing observation around a normal star was the discovery of 51 Peg b with this Doppler method (Mayor and Queloz 1995). Note that the central star is denoted "A" and the planets are labeled with an alphabetical order from "b". The recent discovery of the nearest Earth-sized planet within the habitable zone Proxima b is also made with the Doppler method. This method has detected approximately 740 planets so far. The Doppler method has been most successfully developed at optical wavelengths on many middle- to large-aperture telescopes. However, since nearby M stars are important targets for the coming decades, several high precision near-infrared Doppler instruments such as IRD on the Subaru 8.2-m telescope (Tamura et al. 2012; Kotani et al. 2014) have started their operation or are under development.

28.2.2 Transit Method

The transit method measures small and periodic photometric changes due to the eclipse of the host star by orbiting planets (see Fig. 28.2). Its photometric amplitude is described as:

$$\Delta B / B [\%] \sim 0.01 \left(R_{\text{planet}} [R_{\text{Earth}}] / R_{\text{star}} [R_{\text{Sun}}] \right)^2,$$

where $\Delta B/B$ is the relative brightness change, R_{planet} is the planet radius, R_{star} is the stellar radius, R_{Earth} is the Earth radius, and R_{Sun} is the Sun radius. Photometric accuracies of less than 1% and 0.01% (10 ppm) are necessary to detect Jovian and Earth-like planets, respectively, around Sun-like stars. Note that this method requires a

planetary orbit nearly along the line of sight. The geometric probability of the transit is given by $P = R_{\text{star}}/a$, where R_{star} is the stellar radius and a is the semimajor axis. The transit probability of the solar system planets are 9×10^{-4} and 5×10^{-3} for Jupiter and Earth, respectively. In contrast to the RV method that derives the planet minimum mass, the transit method provides the planet radius. Similar to the Doppler method, the transit method has a bias to large, close-in planets.

There are some significant probabilities of false positives in this method such as grazing stellar binaries, transiting red or brown dwarfs, blended stellar binaries, and others. Therefore, additional observations are required to confirm transit planet candidates. The first exoplanet transit detection was published in 2000 for a hot Jupiter HD 209458 b (Charbonneau et al. 2000; Henry et al. 2000), and approximately 2700 confirmed planets have been detected by this method so far.

The transit method is not only the most successful method to detect exoplanets but also can be applied to characterize the exoplanets. In fact, very accurate observations from space enable the detection of small photometric changes due to the secondary eclipse (when the planet goes behind the central star) and the planet orbital phase (e.g., Deming et al. 2005). These observations can detect thermal emission from planets even without direct imaging and therefore estimate planetary temperatures. Furthermore, transit spectroscopy employing the difference between on- and off-transit measurements enables the characterization of planetary atmosphere, even without direct spectroscopy. The first successful detection of exoplanetary atmosphere using the transit method is the Na D-line detection with the Hubble Space Telescope in HD 209458 (Charbonneau et al. 2002).

In general, the transit method does not provide information on planetary mass. However, in the case of a multiple-planetary transit system, planetary mass can be constrained from the transit timing variation (TTV) as a result of dynamical perturbation among planets. TTV is used to constrain the masses of some recently discovered Earth-like planets around TRAPPIST-1 (Gillon et al. 2017; Grimm et al. 2018; see Sect. 28.4).

28.2.3 *Kepler Mission*

The Kepler telescope was launched in March 2009 (Borucki et al. 2010). It is a NASA Discovery space mission equipped with a 0.95-m Schmidt telescope (1.4-m primary mirror) and 42 optical CCDs covering a field of view of 115 square degrees with a pixel scale of 4 arcsec. From 2009 to 2013 (till its gyro failure occurred), Kepler collected ultrahigh precision photometry of over 190,000 stars simultaneously at a 30-min cadence. After the failure, it is operated as K2 mission (till 2018) to detect exoplanets in 14 fields near the ecliptic plane. The Kepler target stars are relatively faint, ranging from 9 to 16 optical magnitudes. Its achieved photometric accuracy is 30–40 ppm, enough to detect Earth-like planets around Sun-like stars. The Kepler mission has been very successful in detecting exoplanets, resulting in the discovery of approximately 4600 transit planet candidates, of which

approximately 2300 have been confirmed. Most of the Kepler planets are too far to be characterized with the current and near-future telescopes. To detect transiting planets around nearby stars, TESS (Transiting Exoplanet Survey Satellite) is successfully launched in April, 2018, and PLATO (PLANetary Transits and Oscillations of stars) will be in 2026.

28.2.4 Other Indirect Methods

Although many planets have been discovered by the Doppler and transit methods, both are indirect methods, and do not directly image the planets nor distinguish photons from planets and those of host stars. There are several other indirect methods including the microlensing method, the polarimetry method, and the timing method. Among them, the microlensing method has been successful to detect approximately 65 exoplanets, while the pulsar timing method has been successful for detecting some exotic “planets” around neutron stars. However, we do not discuss these in this short contribution.

28.2.5 Direct Imaging and Spectroscopy

Since both the Doppler and transit methods are currently confined to the inner regions of exoplanetary systems, we still know very little about the planets in the outer regions of planetary systems. In addition, because young stars are complicated due to the high level of intrinsic stellar activity, both such surveys have traditionally targeted old and quiet stars. Such Doppler surveys are not suitable for planet searches around massive main-sequence stars due to the paucity of stellar absorption lines and the reduced amplitude of the reflex stellar motion. In contrast to the Doppler and transit methods, direct imaging can be applied to both young and old stars and can allow measurements of colors, luminosities, and spectra, thereby providing mass based on luminosity, temperature, and composition information (see Fig. 28.2). By conducting astrometry based on the direct images, one can also derive kinematical mass.

Although attractive, the direct imaging of exoplanets is an extremely challenging observation. It simultaneously requires (a) high-resolution, (b) high-sensitivity, and (c) high-contrast or high-dynamic range. For example, if one observes our solar system at a distance of 10 pc, (a) the separation between the Sun and Earth is 0.1 arcsec, (b) the apparent brightness of Earth is ~ 30 magnitudes of a star, and (c) the brightness contrast between Sun and Earth is about 10 orders of magnitudes. Although astronomical observations can now achieve such a high resolution and high sensitivity, such a large contrast is beyond the current observation capabilities. However, young low-mass objects including planets are relatively bright, and the contrast between a young host star and planets are smaller than that for an aged

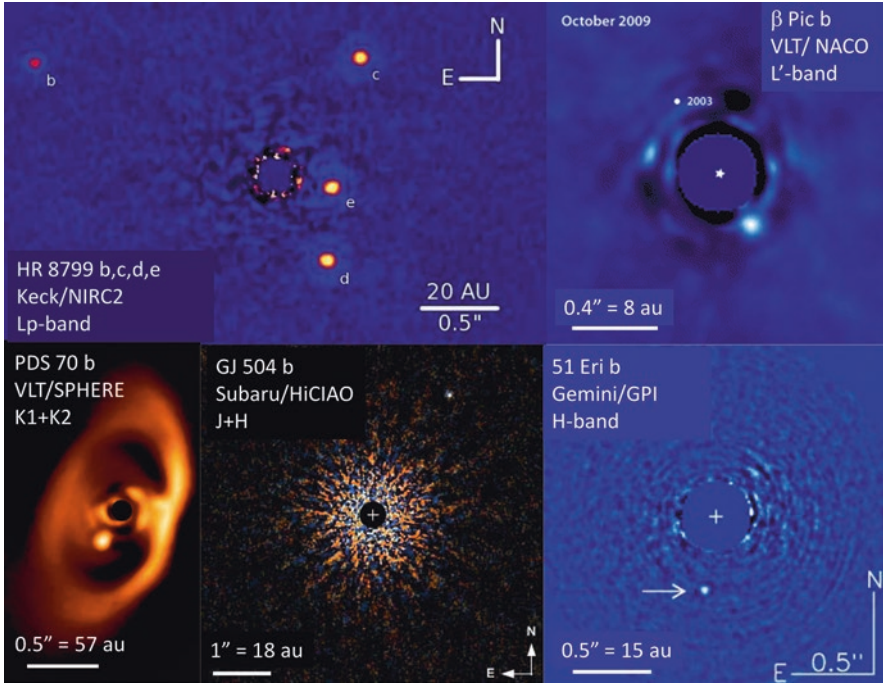


Fig. 28.3 Near-infrared direct images of several self-luminous giant planets obtained with ground-based 8-m class telescopes with adaptive optics. From upper-left in clockwise: HR 8799 b, c, d, e (Marois et al. 2008), β Pic b (Lagrange et al. 2010), 51 Eri b (Macintosh et al. 2015), GJ 504 b (Kuzuhara et al. 2013), PDS 70 b (Keppler et al. 2018), and their host stars are A-type, A-type, F-type, G-type, and T Tauri stars, respectively. Bright light from the central host stars is suppressed by coronagraph and/or differential imaging methods. For β Pic, the planet position in 2003 is also shown. PDS 70 b is situated within the gap of a protoplanetary disk first imaged by Hashimoto et al. (2012)

system. Therefore, current direct imaging targets are self-luminous giant planets around relatively young ($<10^9$ years) host stars (see Fig. 28.3).

For direct imaging of Earth-sized planets around nearest M stars, which requires a high contrast of 10^8 at <0.1 arcsec, the next-generation 30-m class ground-based telescopes such as ELT (Extremely Large Telescope), TMT (Thirty Meter Telescope), and GMT (Giant Magellan Telescope) to be operational in mid- to late 2020s will be necessary. For direct imaging of super-Earths around the nearest Sun-like stars, which requires a higher contrast of 10^9 at a few 0.1 arcsec, the next-generation space mission such as the WFIRST (Wide Field Infrared Survey Telescope) coronagraph planned to be launched in mid-2020s will be necessary. For direct imaging of Earth-sized planets around nearest Sun-like stars or of various kinds of planets around various kinds of stars, future large space telescopes such as HabEx (Habitable Exoplanet Observatory) or LUVOIR (Large UV Optical Infrared) telescope will be necessary.

28.3 Statistics and Characterization of Exoplanets

28.3.1 Planet Occurrence Rate

Demographic of the short-period, inner planets is one of the most important results of the Kepler mission. For all planets with orbital periods less than 50 days, the occurrence rates are $\sim 13\%$, $\sim 2\%$, and $\sim 1\%$ planets per star for planets with radii 2–4, 4–8, and 8–32 R_{Earth} . Therefore, small planets have a very high occurrence rate (Howard et al. 2012).

28.3.2 Habitable Planet Occurrence Rate

The occurrence rates of Earth-sized planets in the circumstellar habitable zones (HZs) in which liquid water could exist on the planet surface are approximately 10% for the Sun-like stars and roughly 50% for M stars, but they vary in the literatures (e.g., see Petigura et al. 2013; Kopparapu 2013). They in fact depend on the definition of HZ and planet size. Note that the most commonly used definition for the HZ is for a planet with the same atmospheric composition and surface pressure as the Earth. The sizes of the HZs are ~ 1 au around G stars and less than ~ 0.1 au for M stars.

28.3.3 Planet Interior Composition

For transiting planets, one can measure both planet radius and mass by the combination of the transit and Doppler methods. The mass-radius relationship allows us to distinguish rocky planets from gas planets as well as rocky terrestrial planets with thin atmospheres from those with thick atmospheres (Fig. 28.4). Most of the planets whose radii below 1.5 R_{Earth} are suggested to have rocky composition (e.g., Weiss and Marcy 2014).

28.3.4 Planetary Atmosphere

Currently there are four methods to observe the atmospheres of exoplanets: (1) transit spectroscopy including multi-wavelength photometry, (2) secondary eclipse thermal emission spectra, (3) high-resolution cross-correlation spectroscopy, and (4) high-contrast direct spectroscopy. (1) Transit spectroscopy involves observing the stellar spectrum both during a transit and outside of a transit and deriving the spectrum of the planet's atmosphere by subtracting these two spectra. Two hot

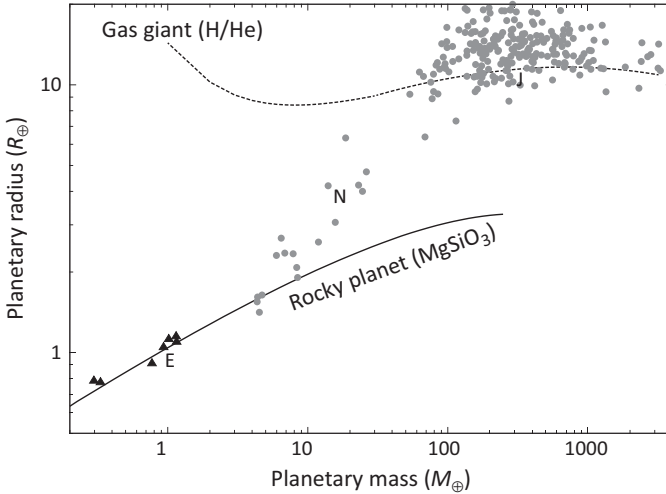


Fig. 28.4 Density distribution of exoplanets in the radius-mass diagram. Only the exoplanets whose planet masses are well determined (signal-to-noise ratios of $M_{\text{planet}} \geq 3$) are plotted. Triangles are seven small planets of TRAPPIST-1. Earth, Neptune, and Jupiter are also shown as the symbols E, N, and J, respectively. The mass-radius curve for rocky planets is from Zeng and Sasselov (2013), and that for gas giant is by courtesy of Dr. Yasunori Hori (coreless pure H/He planets)

Jupiters, HD 209458 and HD189733b, are most well studied so far. Sodium, water vapor, methane, carbon monoxide, and dioxide have been detected. On the other hand, GJ 1214b is the most well-studied super-Earth of $6 M_{\text{Earth}}$ around an M star. The HST (Hubble Space Telescope) and other observations revealed its extremely flat spectra at optical and near-infrared wavelengths and no features in the planet's atmosphere (Kreidberg et al. 2014), indicating the presence of clouds in the atmospheres. (2) The secondary eclipse as the planet moves behind its parent star can be used to determine its temperature and also its composition because we can estimate the thermal emission from the planet. The Spitzer telescope has been most successful in this application (e.g., Demory et al. 2016 for the nearby transiting super-Earth 55 Cancri e). Not only the primary and secondary eclipse but also any phase changes can be traced by very accurate photometry. However, technically it is still limited for hot Jupiters and planets on ultrashort orbits. (3) The high-resolution cross-correlation spectroscopy can extract Doppler-shifted lines of planetary spectra because moving planet lines can be distinguished from stationary telluric and stellar lines (e.g., Snellen et al. 2010). (4) Recent progresses on extreme adaptive optics and IFU (integral field unit) spectroscopy instruments on 8-m class telescopes such as Gemini/GPI, VLT/SPHERE, and Subaru/SCEXAO/CHARIS enable high-contrast direct imaging and spectroscopy. A few spectra covering from 1 to 2.5μ are studied in detail for the wide-orbit planets discovered by direct imaging such as beta Pic b; HR 8799 b, c, d, and e; GJ 504b; and 51 Eri b. Some of them show clear methane, water, and carbon monoxide features.

28.4 Notable Planet for Astrobiology

For astrobiology researches, astronomical observations are essential to provide bio-signature information on exoplanets although they must be performed remotely observed with telescopes. Therefore, the nearest and brightest exoplanetary systems are critically important. Several notable exoplanetary systems are introduced in this section.

28.4.1 *Proxima Centauri and Its Planet b*

Proxima is the nearest star (1.3 pc, 4.2 light year) of a late M-type with a mass of $0.12 M_{\text{Sun}}$ (Anglada-Escudé et al. 2016). It is confirmed to be gravitationally bound to alpha Centauri A ($1.11 M_{\text{Sun}}$) and B ($0.94 M_{\text{Sun}}$). This triple system is composed of a close binary and a wide-orbit companion. Alpha Cen AB binary system orbits each other with semimajor axis of 24 au and eccentricity of 0.524. Proxima has a semi-major axis of 8700 au, eccentricity of 0.5, orbital period of 0.55 Myr, and inclination of 108° with respect to alpha Centauri A and B (Kervella et al. 2017). Proxima b is an Earth-mass ($m_{\text{sin}i}$ of $1.3 M_{\text{Earth}}$) planet receiving 65% of the solar flux (Anglada-Escudé et al. 2016). It is orbiting Proxima with a period of 11.2 days. Based on the orbital properties of Proxima b and its host star, models of the planet's evolution and surface suggest that it could be potentially habitable (e.g., Ribas et al. 2016).

28.4.2 *TRAPPIST-1 and Its Seven Planets b, c, d, e, f, g, and h*

TRAPPIST-1 is an M8-type dwarf at a distance of 12.1 pc (39 light year) with a stellar parameters of $M = 0.089 M_{\text{Sun}}$, $R = 0.121 R_{\text{Sun}}$, and $T_{\text{eff}} = 2500$ K (Van Grootel et al. 2018). The star is faint at optical ($V = 18.8$) but not so much at infrared ($J = 11.4$). It has seven nearly Earth-sized transiting exoplanets (Gillon et al. 2017). The name is after the TRAnsiting Planets and Planetesimals Small Telescope used for its first discovery of three planets. Three of these planets (c, d, e) are within the habitable zone. The small stellar radius enables deep transit depths sufficient for its small planets to be analyzed. All the deep transit depths, the infrared-bright host star, and the frequent transits of a few days to ~ 2 weeks make the TRAPPIST-1 planetary system exceptionally well suited for follow-up infrared transit spectroscopy, especially with the coming JWST telescope. Recent TTV analyses have shown that planets c and e likely have rocky interiors, while planets b, d, f, g, and h require envelopes of volatiles (see Fig. 28.4).

References

- Anglada-Escudé G et al (2016) A terrestrial planet candidate in a temperate orbit around Proxima Centauri. *Nature* 536:437–440
- Borucki WJ et al (2010) Kepler planet-detection mission: introduction and first results. *Science* 327:977–980
- Cassan A et al (2012) One or more bound planets per Milky Way star from microlensing observations. *Nature* 481:167–169
- Charbonneau D et al (2000) Detection of planetary transits across a Sun-like star. *ApJ* 529:L45–L48
- Charbonneau D et al (2002) Detection of an extrasolar planet atmosphere. *ApJ* 568:377–384
- Deming D et al (2005) Infrared radiation from an extrasolar planet. *Nature* 434:740–743
- Demory B-O et al (2016) A map of the large day-night temperature gradient of a super-Earth exoplanet. *Nature* 532:207–209
- Gillon M et al (2017) Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1. *Nature* 542:456–460
- Grimm SL et al (2018) The nature of the TRAPPIST-1 exoplanets. *A&A* 613:A68, 21 pp
- Hashimoto J et al (2012) Polarimetric imaging of large cavity structures in the pre-transitional protoplanetary disk around PDS 70: observations of the disk. *ApJ* 758:L19, 6 pp
- Haswell CA (2010) Transiting exoplanets. Cambridge University Press, Cambridge
- Henry GW et al (2000) A transiting “51 Peg-like” planet. *ApJ* 529:L41–L44
- Howard A et al (2012) Planet occurrence within 0.25 AU of solar-type stars from Kepler. *ApJS* 201:15, 20 pp
- Kasting JF, Whitmire DP, Reynolds RT (1993) Habitable zones around main sequence stars. *Icarus* 101:108–128
- Kepler M et al (2018) Discovery of a planetary-mass companion within the gap of the transition disk around PDS 70. *A&A* 617:A44, 21 pp
- Kervella P et al (2017) Proxima’s orbit around α Centauri. *A&A* 598:L7, 7 pp
- Kopparapu RK (2013) A revised estimate of the occurrence rate of terrestrial planets in the habitable zones around Kepler M-dwarfs. *ApJ* 767:131, 16 pp
- Kotani T et al (2014) Infrared Doppler instrument (IRD) for the Subaru telescope to search for Earth-like planets around nearby M-dwarfs. *Proceedings of the SPIE* 9147. p 9147-914714-12
- Kreidberg L et al (2014) Clouds in the atmosphere of the super-Earth exoplanet GJ1214b. *Nature* 505:69–72
- Kuzuhara M et al (2013) Direct imaging of a cold Jovian exoplanet in orbit around the Sun-like star GJ 504. *ApJ* 774:11, 18 pp
- Lagrange A et al (2010) A giant planet imaged in the disk of the young star β Pictoris. *Science* 329:57–59
- Macintosh B et al (2015) Discovery and spectroscopy of the young Jovian planet 51 Eri b with the Gemini Planet Imager. *Science* 350:64–67
- Marois C et al (2008) Direct imaging of multiple planets orbiting the star HR 8799. *Science* 322:1348–1352
- Mayor M, Queloz D (1995) A Jupiter-mass companion to a solar-type star. *Nature* 378:355–359
- Perryman M (2014) *The exoplanet handbook*. Cambridge University Press, Cambridge
- Petigura EA, Howard AW, Marcy GW (2013) Prevalence of Earth-size planets orbiting Sun-like stars. *Proc Natl Acad Sci* 110:19273–19278
- Ribas I et al (2016) The habitability of Proxima Centauri b. I. Irradiation, rotation and volatile inventory from formation to the present. *A&A* 596:A111
- Seager S (2011) *Exoplanets*. University of Arizona Press, Tucson
- Snellen IAG et al (2010) The orbital motion, absolute mass and high-altitude winds of exoplanet HD209458b. *Nature* 465:1049–1051
- Tamura M (2015) *Exoplanets*. Nippon Hyoron Sha. (in Japanese)
- Tamura M et al (2012) Infrared Doppler instrument for the Subaru Telescope (IRD). *Proceedings of the SPIE* 8446. p 84461T-84461T-10

- Van Grootel V et al (2018) Stellar parameters for Trappist-1. *ApJ* 853:30, 7 pp
- Weiss LM, Marcy GW (2014) The mass-radius relation for 65 exoplanets smaller than 4 Earth radii. *ApJ* 783:L6, 7 pp
- Wolszczan A, Frail DA (1992) A planetary system around the millisecond pulsar PSR1257+12. *Nature* 355:145–147
- Zeng L, Sasselov D (2013) A detailed model grid for solid planets from 0.1 through 100 Earth masses. *PASP* 125:227

Chapter 29

How to Search for Possible Bio-signatures on Earth-Like Planets: Beyond a Pale Blue Dot



Yasushi Suto

Abstract The Earth viewed from outside the Solar System would be identified merely like a pale blue dot, as coined by Carl Sagan. In order to detect possible signatures of the presence of life on a *second Earth* among several terrestrial planets discovered in a habitable zone, one has to develop and establish a methodology to characterize the planet as something beyond a mere pale blue dot. We pay particular attention to the periodic change of the color of the *dot* according to the rotation of the planet. Because of the large-scale inhomogeneous distribution of the planetary surface, the reflected light of the *dot* comprises different color components corresponding to land, ocean, ice, and cloud that cover the surface of the planet. If we decompose the color of the dot into several principle components, in turn, one can identify the presence of the different surface components. Furthermore, the vegetation on the Earth is known to share a remarkable reflection signature; the reflection becomes significantly enhanced at wavelengths longer than 760 nm, which is known as a red-edge of the vegetation. If one can identify the corresponding color signature in a pale blue dot, it can be used as a unique probe to test the presence of life. I will describe the feasibility of the methodology for future space missions and consider the direction toward astrobiology from an astrophysicist's point of view.

Keywords Bio-signatures · Pale blue dot · Red-edge · Copernican Principle

29.1 Introduction

Discovery of an amazing number of exoplanetary systems since 1995 has completely changed our view of the world itself. In particular, we learned once again the universal validity of the Copernican Principle; we do not occupy any special place in the universe. Indeed, this is exactly the very important philosophical lesson that we have learned in the history of astronomy over and over again.

Y. Suto (✉)

Department of Physics and RESCEU (Research Center for the Early Universe),
University of Tokyo, Tokyo, Japan
e-mail: suto@phys.s.u-tokyo.ac.jp

A straightforward corollary of the Copernican Principle is that our Earth is simply just one of the numerous planets in the universe that harbor the life. This will be easily expected from a very crude, order-of-magnitude argument shown below.

The mass of our Galaxy is approximately $10^{11}M_{\text{sun}}$, which implies that there are roughly 10^{11} stars. (In the current argument, we neglect the dark matter contribution and the mass function of stars and simply assume that the typical mass of stars is M_{sun} . This would change the result merely by a few orders of magnitude, and thus the final conclusion below is not affected at all!)

Current planet surveys have revealed that most stars host at least one planet and that dozens out of several thousands of host stars, therefore roughly 0.1%, turn out to have more than one rocky planet located in a habitable zone (e.g., Kasting 1993; Kopparapu et al. 2013), i.e., the equilibrium temperature of the planet is between 0 and 100°C . Thus, if H_2O exists abundantly, it is expected to be liquid on the surface of the planet. The word “habitable” is quite misleading in a sense that the range of equilibrium temperature on the planet surface is supposedly neither a necessary nor sufficient condition for the existence of life. Furthermore, the existence of abundant water on those planets are not at all discovered observationally (yet). Nevertheless, it is a reasonable working hypothesis to proceed further here, and let me use “temperate” instead of “habitable” according to relatively recent literatures.

This implies that we would have 10^8 temperate planets in our Galaxy. It should be emphasized that the value is estimated now from the observed facts. On the other hand, the relative fraction of planets, p_{water} , with a reasonable amount of water, and possibly with a reasonable amount of lands on the surface as well, is quite uncertain at this point but will be estimated observationally in the future by a remote sensing as described in Sect. 29.3.

If a planet has a right amount of water for bearing life, what is the probability that the planet eventually develops life? This is intrinsically difficult and almost impossible question to answer scientifically. Therefore, we need to resort to the Copernican Principle. Our Solar System was born about 4.6 Gyr ago, and the first life on the Earth is supposed to have emerged approximately 1 Gyr later. Thus, the emergence of life itself may not be such a rare event as long as the relevant environment, which we do not yet understand exactly, is provided. A simple application of the Copernican Principle suggests that a fairly large fraction of temperate planets with oceans and lands will inevitably develop a certain type of life.

Plants on the Earth went out of oceans and started to grow on lands about 0.5 Gyr ago; from the viewpoint of remote sensing, this is the most important event in the evolution of life. It is not clear at all how long such planets continue to be detectable via remote sensing, since we have no idea if any life-form exhibiting a significant bio-signature survives possible drastic environment changes including astronomical impacts, geophysical activities, and human wars. Even if we assume pessimistically that our Earth stops exhibiting detectable bio-signatures very soon, the fraction of detectable period of our Earth via remote sensing (of course, just in principle) over the lifetime of the Sun would be $0.5 \text{ Gyr}/10 \text{ Gyr} = 0.05$.

Therefore, out of the 10^8 temperate planets in our Galaxy, $5 \times 10^6 p_{\text{water}}$ would potentially exhibit some kinds of bio-signatures for remote-sensing. Again, on the basis of the Copernican Principle, p_{water} is unlikely to be sufficiently small, and we

expect that $5 \times 10^6 p_{\text{water}} \gg 1$. This implies that we have to consider seriously how to search for possible bio-signatures on Earth-like planets.

29.2 Lessons from Previous Pioneering Attempts

Vesto Melvin Slipher is a renowned astronomer, well known for his contribution to the discovery of redshifts of distant galaxies. Indeed, Edwin Hubble owed Slipher's measurements of galaxy redshifts in proposing his "famous" distance-redshift relation (Hubble 1929), which eventually led to the standard model of the expanding universe. It is now well recognized that Slipher's contribution to the discovery of the expansion of the universe has been significantly underestimated; I would definitely recommend Peacock (2013) for interested readers, which nicely describes numerous great achievements of Slipher.

Actually, his pioneering contribution to astrobiology seems to have been equally underestimated either. In his paper entitled "Observations of Mars in 1924 Made at Lowell Observatory II. Spectrum observations of Mars" (Slipher 1924), he attempted to test the existence of chlorophyll in the dark region on Mars. He clearly recognized the importance of the reflection spectrum feature of vegetation as a possible bio-signature on Mars. He noted that "The reflection spectrum from vegetation is not at all definite visually as its most distinctive feature is its brilliancy in the deep red, beyond the sensitivity of the eye."

This characteristic feature, often referred to as the red-edge of vegetation (sharp increase of reflection spectrum beyond around $0.75 \mu\text{m}$), is supposed to be very generic over most plants on Earth and understood to be related to the efficiency of the photosynthesis.

He concluded his paper by stating that "The Martian spectra of the dark regions so far do not give any certain evidence of the typical reflection spectrum of chlorophyll. The amount and types of vegetation required to make the effect noticeable is being investigated by suitable terrestrial exposures." I believe that this is quite amazing and pioneering work in the history of astrobiology.

Interestingly, there have been several observational claims of spectroscopic evidence for vegetation on Mars on the basis of a different absorption feature around $3.4 \mu\text{m}$ (e.g., Sinton (1957) and also Briot et al. (2004) for a historical overview). Assuming that they are not reliable, Slipher's idea has been seriously considered and applied in the modern context for the first time by Sagan et al. (1993). They searched for bio-signatures on Earth at the first flyby of Galileo spacecraft on December 8, 1990, and successfully concluded that there is life on Earth!

Bio-signatures that they detected from the remote sensing of Earth include (i) abundant gaseous oxygen in visible and near-infrared bands, (ii) atmospheric methane of the significantly larger abundance than expected from simple thermal equilibrium, and (iii) red-edge feature around $0.75 \mu\text{m}$.

Figure 29.1 plots the spectra of Earth over a relatively cloud-free region of the Pacific Ocean observed by Galileo (reprinted from Fig. 1 in Sagan et al. 1993). Figure 29.1a is the spectrum for $0.70 < \lambda[\mu\text{m}] < 1.0$, showing a strong absorption

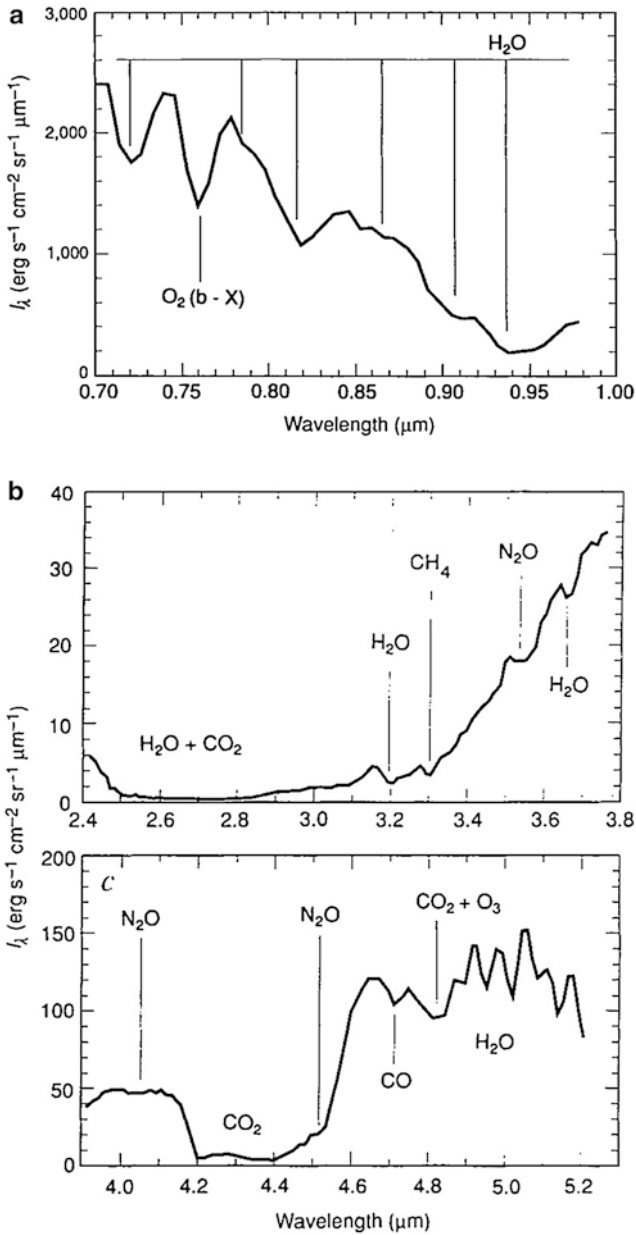


Fig. 29.1 Spectra of Earth observed by Galileo spacecraft in December 1990 over a relatively cloud-free region of the Pacific Ocean. The molecules responsible for major absorption bands are indicated in each panel. (Reprinted from Fig. 1a–c of Sagan et al. 1993)

feature of the A band molecular oxygen at $0.76 \mu\text{m}$ with several H_2O absorptions as well. The column density of O_2 is estimated to be about 200 g cm^{-2} , and such a large abundance is very unlikely to be produced and accumulated by any abiotic process like the UV photodissociation of water followed by the Jeans escape of hydrogen to space. Thus Sagan et al. (1993) concluded that “Galileo’s observations of O_2 , thus at least raise our suspicions about the presence of life.”

In the near-infrared bands, $2.4 < \lambda[\mu\text{m}] < 3.8$ (Fig. 29.1b) and $3.9 < \lambda[\mu\text{m}] < 5.3$ (Fig. 29.1b), a very strong CO_2 absorption band can be clearly identified around $4.3 \mu\text{m}$, as well as several N_2O and H_2O absorptions. Most notably, they identified the methane feature at $3.31 \mu\text{m}$ and found that the derived abundance is about 140 orders of magnitude higher than a simple expectation from thermal equilibrium. Since CH_4 is supposed to be oxidized quickly to CO_2 and H_2O , a continuous pumping source of CH_4 is required, which is most likely life. This is also the case for the high disequilibrium abundance of N_2O , which will be due to the presence of nitrogen-fixing bacteria and algae.

Such abundant atmospheric molecules can be used as important bio-signatures in classical astronomical observations, but still may not be directly related to the presence of life. Indeed the biological interpretation of the observed methane and other molecules may not be so robust; for instance, Epiope and Sherwood Lollar (2013) discussed the possible abiotic origin of methane in the planetary atmosphere. The detection of the red-edge feature, in turn, is challenging but, if detected at all, would be interpreted as a more straightforward evidence for the life dominating a fair fraction of the planet.

As shown in Fig. 29.2 (reprinted from Figs. 2c and 3 in Sagan et al. 1993), Sagan et al. (1993) presented broadband spectra of the three different regions on Earth and

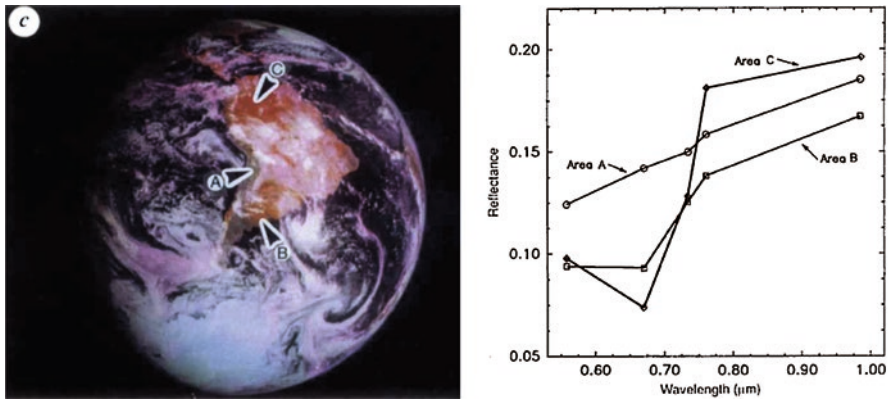


Fig. 29.2 (Left) Composite color image of Earth observed by Galileo spacecraft in December 1990. (Right) Spectra corresponds to the three areas A, B, and C in the left panel. (Reprinted from Figs. 2c and 3 of Sagan et al. 1993)

argued that the unusually strong absorption features in the spectra of areas B and C are “the signature of a light-harvesting pigment in a photosynthetic system,” while that of area A is consistent with a variety of dark rock or mineral-soil surfaces.

29.3 Colors of a Second Earth

It is no doubt that Sagan et al. (1993) is the first serious observational attempt to present fundamental methodologies to search for life in planets from remote-sensing data. It is even more amazing to recognize that it was before the first discovery of an exoplanet around a sun-like star (Mayor and Queloz 1995). Unfortunately, however, their method was partially based on the spatially resolved imaging observation of the surface of Earth, and thus is not directly applicable to exoplanets.

Ford et al. (2001) is the first to realize such a basic limitation. They computed diurnal photometric variability of Earth in different bands by averaging over the visible part illuminated by the Sun for a distant observer. In particular, they claimed that the photometric variability due to the red-edge feature may be marginally detectable for future space interferometer missions. The diurnal photometric variability is a more realistic approach with remote-sensing of Earth-like exoplanets since it is based on the continuous monitoring of a change of colors of spatially “one dot.” Indeed, this implies that Earth is not a mere *pale blue dot* but a color-changing dot due to its spin-modulated surface landscape.

Inspired by this basic idea, we started a systematic study of the diurnal photometric variability of Earth (Fujii et al. 2010, 2011). We have not only computed the expected variability but also attempted to estimate the area fraction of different surface components including snow, land, ocean, and vegetation by inverting the simulated light curves in different photometric bands (see also Cowan et al. 2009; Majeau et al. 2012; Fujii et al. 2017, and references therein).

To be more specific, we first created mock light curves for Earth *without clouds* in different photometric bands, using empirical data from satellites. These light curves were attempted to be fit to an isotropic scattering model consisting of four surface types, ocean, soil, snow, and vegetation, as shown in Fig. 29.3. We considered a very idealized observational situation in which the light from the host star is completely blocked and the photometric noise is due to the Poisson fluctuations in the observed photon counts from the planet alone.

Figure 29.4 presents an example of our decomposition of light curves in terms of the four surface components using the photometric bands (1)–(5) indicated as gray bars in Fig. 29.3. This simulated observation assumes an Earth-twin from 10 pc away from us and a dedicated space telescope of diameter 2 m with an exposure time of 1 h over 2-week continuous monitoring. In such an idealized situation where the light from the host star is completely blocked and the planetary surface is not covered by clouds, we are able to recover the correct fractional areas of surface components fairly well. In particular, at a certain phase of Earth in which plants

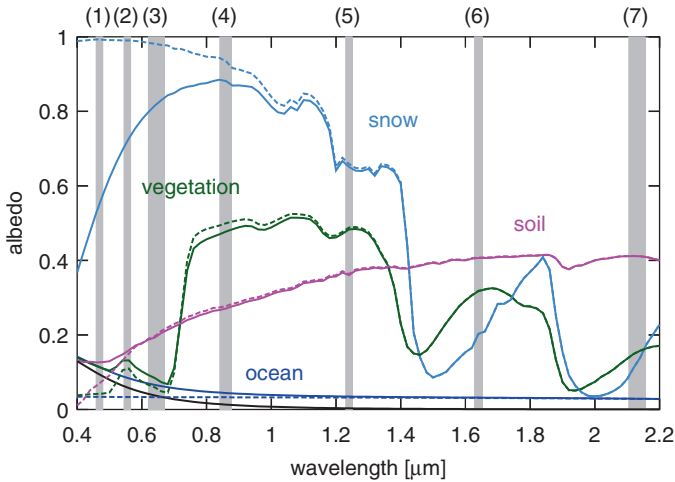


Fig. 29.3 Wavelength-dependent effective albedos of ocean (blue), soil (magenta), vegetation (green), snow (cyan), and atmosphere with Rayleigh scattering alone (black). The solid lines show the effective albedo with Earth-like atmosphere, while the dashed lines show the effective albedo without an atmosphere. Shaded regions correspond to the MODIS (Moderate Resolution Imaging Spectroradiometer) bands. The numbers at the top are the labels of the different photometric bands. (Reprinted from Fig. 7 of Fujii et al. 2010)

cover a large fraction of the planetary surface, we may be able to even detect the presence of vegetation via its distinct spectral feature of photosynthesis.

Admittedly our assumptions may not be so realistic. Especially, the significant cloud coverage would be inevitable for any temperate terrestrial planets with abundant liquid water.

Therefore, we considered the degree of degradation of our surface recovery method due to the cloud coverage by applying it to multi-band diurnal light curves of Earth from the EPOXI spacecraft. The detailed discussion can be found in Fujii et al. (2011), and here we simply show the resulting longitudinal map recovered from the diurnal variations in Fig. 29.5. While the angular resolution is inevitably low, it is encouraging that some of the major geographical features of the Earth, e.g., two oceans, the Sahara desert, and the two largest land masses, can be approximately identified, even from the color-changing dot alone.

29.4 Conclusion

Apparently, there remain many things to be improved in methodology both theoretically and observationally. Fujii et al. (2010, 2011) exploited the color modulation of the visible surface due to the planetary spin alone. If the planetary spin axis is oblique with respect to its orbital axis, as is the case with Earth, the resulting annual

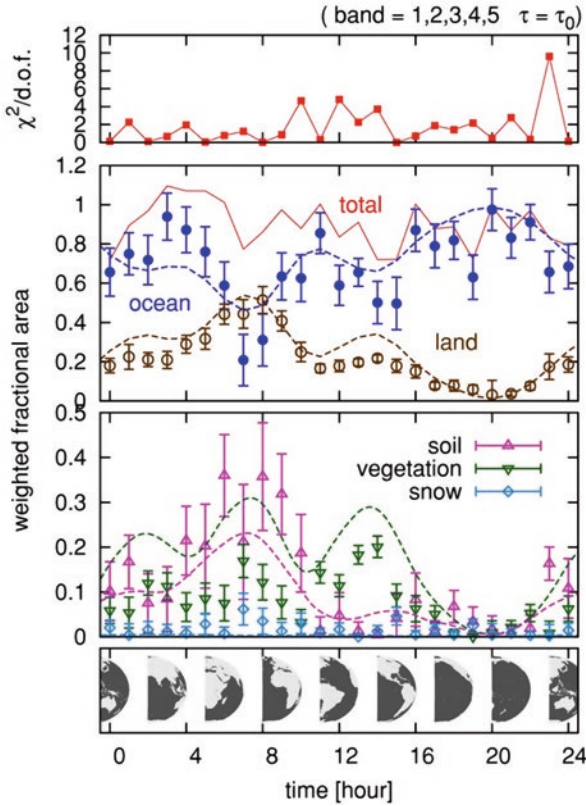


Fig. 29.4 Reconstructed fractional areas for four surface types from the simulated light curves in five bands (bands 1–5). The top panel shows the value of reduced χ^2 for each epoch. The upper middle panel displays the results of estimating weighted fractional areas of ocean (blue), land (=soil+vegetation+snow; brown), and the total of them (red). The lower middle panel displays those of soil (magenta), vegetation (green), and snow (cyan). The dashed lines in those two panels show the weighted fractional areas derived from the real classification dataset by the MODIS satellite. The quoted error bars indicate the variance of the best-fit values from 100 realizations. The bottom panel depicts the snapshots of the Earth at the corresponding epochs where the ocean is painted in gray and the land in white. (Reprinted from Fig. 8 of Fujii et al. 2010)

modulation may even allow the recovery of the two-dimensional surface map of the planet. This interesting possibility has been studied by Kawahara and Fujii (2010, 2011) and Fujii and Kawahara (2012), and they showed that it is possible to reproduce the 2D surface map of Earth, at least in an idealized situation. Kawahara (2016) also proposed novel methodology to determine the planetary obliquity from frequency modulations of photometric light curves.

Those preliminary feasibility studies have adopted simulated and/or observed datasets of Earth itself. Needless to say, it is by far the most important check to begin with, but it is also true that such studies never cover a possible range of realistic diversities of candidate planets. This is a fundamental and intrinsic limitation in

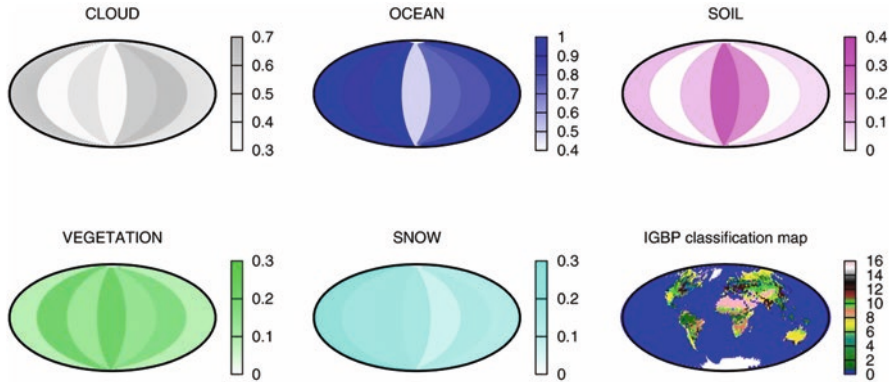


Fig. 29.5 The 7-slice longitudinal distribution of each surface component recovered from the June light curves. The indices of the IGBP classification map (the bottom right panel: <http://modis-atmos.gsfc.nasa.gov/ECOSYSTEM/index.html>) are 0, ocean; 1, evergreen needleleaf forest; 2, evergreen broadleaf forest; 3, deciduous needleleaf forest; 4, deciduous broadleaf forest; 5, mixed forest; 6, closed shrubland; 7, open shrubland; 8, woody savannas; 9, savannas; 10, grasslands; 11, permanent wetlands; 12, croplands; 13, urban and built-up; 14, cropland/natural vegetation mosaic; 15, snow and ice; and 16, barren or sparsely vegetated. (Reprinted from Fig. 16 of Fujii et al. 2011)

considering bio-signatures, given the fact that we have no idea of life outside Earth. Nevertheless, it is such an important and fascinating topic, not only in astronomy and astrobiology but even in all modern sciences. We need to continue and definitely will be able to realize that “We did not know anything” (Asimov 1941).

References

Asimov I (1941) Nightfall in Astounding Science Fiction magazine
 Briot D, Schneider J, Arnold L (2004) G.A. Tikhov, and the beginnings of astrobiology. In: Beaulieu J-P, Lecavelier des Etangs A, Terquem C (eds) Proceedings of “Extrasolar planets: Today and Tomorrow”. ASP Conference Series 321. pp 219–220
 Cowan NB et al (2009) Alien maps of an ocean-bearing world. *Astrophys J* 700:915–923
 Etiope G, Sherwood Lollar B (2013) Abiotic methane on earth. *Rev Geophys* 51:276–299
 Ford E, Seager S, Turner EL (2001) Characterization of extrasolar terrestrial planets from diurnal photometric variability. *Nature* 412:885–887
 Fujii Y, Kawahara H (2012) Mapping earth analogs from photometric variability: spin-orbit tomography for planets in inclined orbits. *Astrophys J* 755:101 (pp.1–14)
 Fujii Y, Kawahara H, Suto Y, Taruya A, Fukuda S, Nakajima T, Turner EL (2010) Colors of a second earth: estimating the fractional areas of ocean, land, and vegetation of earth-like exoplanets. *Astrophys J* 715:866–880
 Fujii Y, Kawahara H, Suto Y, Fukuda S, Nakajima T, Livengood TA, Turner EL (2011) Colors of a second Earth. II. Effects of clouds on photometric characterization of Earth-like exoplanets. *Astrophys J* 738:184 (pp.1–15)
 Fujii Y, Lusting-Yaeger J, Cowan NB (2017) Rotational spectral unmixing of exoplanets: degeneracies between surface colors and geography. *Astron J* 154:189 (pp.1–9)

- Hubble E (1929) A relation between distance and radial velocity among extra-galactic nebulae. *Proc Natl Acad Sci* 15:168–173
- Kasting JF (1993) Earth's early atmosphere. *Science* 259:920–926
- Kawahara H (2016) Frequency modulation of directly imaged exoplanets: geometric effect as a probe of planetary obliquity. *Astrophys J* 822:112 (pp.1–11)
- Kawahara H, Fujii Y (2010) Global mapping of earth-like exoplanets from scattered light curves. *Astrophys J* 720:1333–1350
- Kawahara H, Fujii Y (2011) Mapping clouds and terrain of Earth-like planets from photometric variability: demonstration with planets in face-on orbits. *Astrophys J (Lett)* 739:L62 (pp.1–6)
- Kopparapu RK et al (2013) Habitable zones around main-sequence stars: new estimates. *Astrophys J* 765:131 (pp.1–16)
- Majeau C, Agol E, Cowan NB (2012) A two-dimensional infrared map of the extrasolar planet HD 189733b. *Astrophys J* 747:L20 (pp.1–5)
- Mayor M, Queloz D (1995) A Jupiter-mass companion to a solar-type star. *Nature* 378:355–359
- Peacock JA (2013) Slipher, galaxies, and cosmological velocity fields, way. In: Way MJ, Hunter D (eds) Proceedings of “Origins of the Expanding Universe: 1912–1932”, APS conference series 471. pp 3–23
- Sagan C, Thompson WR, Carlson R, Gurnett D, Hord C (1993) A search for life on Earth from the Galileo spacecraft. *Nature* 365:715–721
- Sinton WM (1957) Spectroscopic evidence for vegetation on Mars. *Astrophys J* 126:231–239
- Slipher VM (1924) Observations of Mars in 1924 made at the Lowell Observatory II Spectrum observations of Mars. *Astron Soc Pac* 36:261–262

Chapter 30

SETI (Search for Extraterrestrial Intelligence)



Hisashi Hirabayashi

Abstract The technology to communicate over interstellar distances is expected to be developed for civilization anywhere in the universe. In this chapter, the strategy and the activities for searching for extraterrestrial intelligence are reviewed. The search started by receiving radio wave in 1960 and has been continued with a modest activity. Optical communication can be another target, and optical searches have also been started. Because of the continuing technological leaps, the increasing possibility in finding extraterrestrial intelligence is expected. To find the better way to search extraterrestrial intelligence successfully, we need to understand our universe more.

Keywords SETI · Interstellar communication · Intelligent life · Galactic civilization

30.1 Introduction

Is there intelligent life beyond the Earth? Though some ancient Greek philosophers considered the possibility, there was no means to undertake a scientific investigation. Though it is still impossible to travel over interstellar distances, it is possible to scientifically approach this fundamental question.

It was an important step that the physicists Cocconi and Morrison (1959) discussed the idea of interstellar communication using radio wave. By that time, large radio telescopes had been built, and the space communication era was emerging. As the 21 cm (1420 MHz) radio line of interstellar hydrogen atoms had been detected (Ewen and Purcell 1951), it was proposed to be the interstellar communication band because no other line had been detected. To circumvent a huge unknown parameter space in frequency, some adequate assumptions had to be made which frequency to be searched.

H. Hirabayashi (✉)

The Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency,
Sagamihara, Kanagawa, Japan
e-mail: hirax@tbr.t-com.ne.jp

F. Drake of the US National Radio Astronomy Observatory (NRAO) conducted the first search near the hydrogen line frequency for two nearby sun-like stars, Tau Ceti and Epsilon Eridani, both about 11 light-years in distance from our solar system in Project Ozma. He observed them for a total of 200 h over 2 months in 1960. Though he did not detect any sign of an intelligent signal, it triggered a number of other searches in the following years. SETI is now the technical term representing the search for extraterrestrial intelligence, and academic activity has been authorized in the working groups in the IAU (International Astronomical Union) and the IAF (International Aeronautical Federation). So far, there have been more than 100 searches made over almost 60 years, with no firm detection. However, the fact does not prove that there is no extraterrestrial intelligence: the absence of evidence is not evidence of absence.

This chapter reviews how SETI has been conducted and what are the future prospects.

30.2 Radio Searches and Search Strategy

There is a vast parameter space in which search for unknown civilizations is to be conducted: direction, time, frequency, bandwidth, and so on. The combination of these uncertain factors quickly piles up to a huge number – searching for the signal is much more difficult than establishing human communications. As a result, most searches to date have done in a very limited parameter space. The search instrument must have a huge parallel processing capability, and it is crucial making careful choices in this parameter space.

In terms of the target directions, the three main types of searches are present:

- All-sky: An unbiased sky survey, scanning all the accessible sky.
- Targeted: Directed toward selected stars.
- Serendip-type: Searches by “piggy-backing” on other radio astronomy observations by independently processing the data. In this case, astronomically interesting sources are the targets.

It is worth mentioning that groups at both Ohio State University and Harvard University, USA, maintained dedicated continuous searches using their own radio telescopes for many years. The Ohio group started the search in 1973. Though they announced that they detected a strong narrowband signal near the hydrogen line on August 15, 1977, called the “Wow! signal,” it has never been detected again. Recently Paris and Davies (2017) reported that hydrogen clouds from comets 266P Christensen and P2008 Y2 (Gibbs) are candidates for the source of the 1977 Wow! signal.

Over the years, the search capabilities of telescopes have increased greatly in terms of antenna design, receiver sensitivity, data processing power, etc.

The frequency band that offers the best signal-to-noise ratio is the microwave region. At lower frequencies, our galaxy is very bright, due to synchrotron radiation

produced by fast electrons in the galactic magnetic field. The 1–10 GHz range is good for ground-based telescopes, as it avoids atmospheric emission and absorption and the effects of receiver quantum noise at higher frequencies. However, we might expect that advanced civilizations can easily build space-based facilities, and the atmosphere may not be a major limitation. Therefore, 1–100 GHz is thought to be the proper frequency band to search.

The antenna and receiver need to be sensitive to a wide frequency band, which is processed by a spectrometer to search for a narrowband signal. Digital processing is commonly used for this purpose. For a successful detection over interstellar distances, the signal must be concentrated in a very narrowband, say 0.1–1 Hz, which is limited by interstellar scattering and scintillation (analogous to the twinkling of starlight). Spectral lines from natural phenomena are generally much broader than this. Thus, the search parameter of frequency channel is the order of $100 \text{ GHz}/0.1 \text{ Hz}$. When viewed by large antennas in the microwave band (where they have an angular resolution similar to human eyesight, roughly 1 min of arc), the whole sky has the order of 10^{11} independent directions. Therefore, the combination of directional and frequency channel uncertainties at the time of the search is in the order of 10^{22} , which is close to Avogadro's number! The number of stars in the galaxy and the number of galaxies in the observable universe are both roughly 10^{11} . It is interesting to consider these numbers and to appreciate the vast unknown parameter space for SETI.

Obtaining a spectrum at the 1 Hz level resolution for each observational direction takes a large amount of computational time. D. Werthimer of UCB invented and started a smart scheme of using innumerable remote PCs, which otherwise are in screen saver mode, to automatically download the processing software, download the necessary data stream, process the data, and to send back the result (SETI@Home). This is called “grid computing,” and it is now applied in many other fields. This is encouraging not only for workload sharing but also making PC owners feel that they have joined and are really contributing to the project.

There have been several major searches worth mentioning. NASA planned and started a big SETI project in 1992 but soon canceled it. A group of researchers including F. Drake and J. Tarter founded the SETI Institute and effectively carried out the NASA project with the name of Project Phoenix, relying on donations from private sources. Between 1995 and 2004, Project Phoenix targeted 800 stars up to 240 light-years away with the world's largest radio telescopes over a frequency range of 1.2–3 GHz. The dedicated BETA project performed an all-sky, narrowband microwave search between 1400 and 1720 MHz with billions of frequency channels with the dedicated use of a 26 m radio telescope. This band covers hydrogen (H, 1420 MHz) and hydroxyl (OH, 1666 MHz) line frequencies at the two ends of the band.

For more detail, see the review paper written by Tarter (2001).

30.3 Future Prospects of Radio SETI

Radio telescopes have improved in sensitivity, field of view, spectral and time resolution, etc. An array of a large number of antennas with small apertures, with proper later processing, has a wide field of view, good angular resolution. This technique, called “aperture synthesis,” was invented by the Cambridge radio astronomer, Sir M. Ryle, for which he shared the 1974 Nobel Prize in Physics (Ryle 1962). In this technique, an array of radio telescopes can be regarded as many combinations of Michelson interferometers which obtain two-dimensional Fourier components of the radio source image. In this technique, the source image can be digitally reconstructed by two-dimensional Fourier inverse transform. Spectral analysis relies also on digital processing, mostly on the fast Fourier transform (FFT), and has been advanced following Moore’s law, which states that processing power doubles approximately every 18 months.

The Allen Telescope Array (ATA) of UC Berkeley and the SETI Institute was designed for both radio astronomy and SETI, using the aperture synthesis technique. It started with 42 antennas with 6 m diameter, having plans (which were never realized) to increase to 350 antennas, and performed SETI searches. This was the first time that SETI was considered from radio telescope design phase, and the funds were donated from P. Allen, a co-founder of Microsoft.

A far more powerful radio telescope of this type is the international Square Kilometer Array (SKA) which is expected to be operational in the mid- to late 2020s. This will be the most powerful radio telescope for radio astronomy and SETI. The SKA project, led by ten member countries, will have two telescope sites, South Africa and West Australia. In Phase 1, an array of dishes in South Africa will be used at higher frequencies, and an array of dipole antennas in Western Australia will be used at lower frequencies. The plans for a later Phase 2 would extend telescopes across Africa and antennas across Australia. This would be a very big leap for SETI with the SKA. Military radars on planets within 300 light-years from Earth could be detectable by the SKA.

Breakthrough Listen is a SETI program announced in 2016 by the board members Y. Milner (founder of DST Global, who is funding the project) and S. Hawking and is expected to be considerably bigger than Project Phoenix. The same group started a design study to send very small probes to the newly discovered nearest exoplanet Proxima Centauri b at a distance of 4.26 light-years. They do not rely on a chemical or nuclear energy-powered rocket but on photon beam pressure from an Earth-based high-power laser array. Thousands of tiny probes are planned to be sent in 20 years at a 20% of the light speed to fly by and to image the planet.

30.4 Possible Types of Signal

It is important to consider the possible types of SETI signal. Three types of signals should be considered separately in the SETI activity.

- (a) **Incidental signal:** This includes radars, broadcasting signals, and communication signals, which are emitted from normal activities within an extraterrestrial civilization without intention for interstellar communication. Earth has been transmitting such signals for about a hundred years.
- (b) **Intentional signals:** The signals emitted from extraterrestrial intelligence aiming the communication to other intelligent presence.
- (c) **The signal for transmitting information flow with modest bandwidth.** This is the communication signal after the first radio contact is established.

If SETI succeeds in finding another intelligence, then it can be naturally expected to have second and third discoveries of extraterrestrial intelligence. Assuming that we are not the intelligent presence developed in earlier phase, there may be more advanced civilizations having communication network each other with the signal of type c) above. R. Bracewell referred to this network as the “Galactic Club.”

30.5 Optical SETI

Laser-based interstellar communication was proposed by C. Townes, one of the inventors of the laser. Laser power has increased significantly in recent years. If another civilization has developed the laser with similar power, and a high-power laser is connected to a 10 m class antenna, the signal from a planet up to 100 light-years away could outshine the star it is orbiting. For this reason, civilizations may emit high-power optical pulses as an active signal for SETI communication purpose. P. Horowitz of Harvard University, who has been very active in radio band SETI, started optical SETI with a 72-in. telescope at the Harvard-Smithsonian Observatory, to try to detect light pulses with a time resolution of 1 ns (Howard et al. 2000). The targeted all-sky Optical SETI project pointed at 6000 star systems. With the development of multi-pixel detectors at the focus, the search capability can be increased significantly. An advantage of optical SETI is that pulsed emission will not suffer from smearing by the dispersion effect by interstellar plasma. On the other hand, at radio wavelengths the dispersion effect is large, and sophisticated frequency analysis is needed.

30.6 The Number of Galactic Civilizations

In 1961, a SETI meeting was held at the Green Bank Observatory of US National Radio Astronomy Observatory, and F. Drake showed a formula which is now called the Drake equation (Drake and Sobel 1992). This is a very simple formula to estimate the number N of civilizations in the galaxy.

$$N = RfpneflifcL$$

The number N is obtained by multiplying R , the number of newborn stars per year in the galaxy, by L , the mean lifetime of civilization in year, and by several fractional factors. These factors are the following: fp = the fraction of stars with planets, ne = number of Earth-like planets per stellar system, fl = fraction of planets with life, fi = the fraction of planets with intelligent life, and fc = the fraction of planets with communication technology. The equation combines terms related to astronomy, planetary physics, biology, and intelligence. The number of stars being formed in the galaxy is an astronomical question and is known to be about one to ten stars per year.

Following the successful detection of thousands of extrasolar planets since 1995, we are gaining a good understanding of the probability of planet formation. The fraction of having Earth-like planets is also improving.

However, there are no reliable values for factors fl , fi , and fc . The factors may be 1 from the very optimistic point of view, while the factors may be 0.01 or less from the very pessimistic point of view. There is basic difference between the two views, natural or miraculous nature of evolution. Some discussions and speculations can be found in Ulmschneider (2003), for example. It is also known that the changes and events of Earth environment have significantly influenced the evolution of life and intelligence. The better understanding of life and intelligence may give better convergence and estimate of the factors in the future.

30.7 Nature of Advanced Civilizations and Artificial Intelligence

The final term L raises a far more difficult question. How long do civilizations last? Or what is the mean lifetime of galactic civilizations? One can imagine astronomical scale lifetime of, say, 1 giga-year, assuming stable civilization. On the other hand, 1000 years or even less may be anticipated from pessimistic point of view for our civilization. It is difficult to extrapolate and imagine wise and mature galactic civilizations. However, we would need very long years of mean lifetime in order to have a chance of exchanging information with potential partners in the galaxy.

Civilizations must develop sufficient stability to ensure not to destroy themselves nor to become extinct. Here we may call stable and long-lived civilizations as

“super-civilizations.” Earth’s current civilization has many problems to solve and to survive, and it cannot be regarded as a super-civilization yet. We do not know whether super-civilizations are interested in other worlds or focused on internal and imaginary worlds. This is probably the biggest uncertainty in estimating the number of galactic civilizations. Only after SETI being successful can we be sure that super-civilization is not an illusion and start to know the mean lifetime or probability of super-civilization. Extraterrestrial intelligence can be found only if these factors are large enough.

Though the type of individual members of those civilization is not known, it is interesting to imagine them. They may differ from their biological original form of life, being replaced by artificial intelligence. There is a simple forecast that artificial intelligence (AI) will surpass humans around 2045, as portrayed in the movie “Singularity.” Bernal (1929) discussed human, world, and a future civilization where all members are connected by information technology to form a united structure. Though Earth’s civilization is becoming connected by the Internet, it is not a coherently linked structure. It is worth imaging deeper and further the future connected mode of individuals.

30.8 Astronomical Research and SETI

Pulsars were discovered serendipitously when the Cambridge group started observing scintillating radio sources in 1967 (Hewish et al. 1968). The precise pulse trains of pulsars were at first thought to be the sign of an artificial signal. However, the pulses were soon understood as the beamed radiation from the magnetosphere of rapidly spinning neutron stars (Gold 1968).

Fast radio bursts (FRB) were first found in 2007 (Lorimer et al. 2007) as a single strong pulse event with very short duration, about 1 ms. There are clear signs of significant intergalactic dispersion, as the pulses show a “chirp” nature (the higher frequency pulse part comes first, and the lower frequency part comes later). Therefore, the signals must almost certainly come from outside our galaxy. By the short duration of the pulse, the emission region must be less than the light travel time of 1 ms, or 30 km. Though the origin of these pulses is not understood yet, it must be due to exotic and energetic stellar phenomenon. An important discovery came in 2016 when it was found that the FRB121102 event was followed by repeated pulse events (Paris and Davies 2017). By later precise radio measurement of the source direction, it was found that the events come from the direction of faint dwarf galaxy three billion light-years away, though the origin of the emission is still not understood.

We do not know all the types of exotic phenomena in the universe yet. We do not know the status of advanced galactic civilization nor can predict the possible characteristics of their signals properly. In astronomy, we do not understand the constituents of dark matter, and do not understand the nature of dark energy, even though they are now dominant in the universe. However, our understanding has

increased greatly since the 1960s when we first started to answer basic questions about the universe. We should continue these efforts to observe the universe, try to understand our physical world, and must keep our minds open. This is important also for SETI.

We live on the mother Earth with most of our energy derived from our sun. We think this is our comfortable and sustainable environment. However, some advanced civilizations may derive their energy from much more energetic astronomical objects, a spinning neutron star, a black hole, or even more powerful phenomena. N. Kardashev classified galactic civilizations into three categories based on their level of energy consumption (Kardashev 1964): type I civilization which can use all of the planetary energy, type II civilization which can use all of the stellar energy, and type III civilization which can use all of the galactic energy. However, we do not know whether super-civilizations rely on very energetic phenomena or rather make economical and smart use of energy.

SETI is a long-term endeavor of mankind. From the arguments above, the author's suggested guidelines for SETI are:

- Full efforts should be continued to keep Earth's civilization.
- Continue SETI activities at a proper level.
- Continue astrophysical research to understand our universe.

30.9 Conclusion

SETI started in the 1960s. SETI was initially conducted in the radio band, with optical band searches joining later. Though about 100 radio searches have been conducted until now, there is no convincing detection of an artificial extraterrestrial signal yet. There is big unknown parameter space to be searched, and further efforts should be continued. The search capability of instruments is still developing rapidly, and no one can predict when SETI may be successful. We do not know if totally new phenomena connected with super-civilizations' activity are to be discovered. We must be open-minded.

References

- Bernal JD (1929) *The world, the flesh and the devil: an enquiry into the future of the three enemies of the rational soul*, Cape edn. Jonathan Cape, London
- Bracewell RN (1974) *The galactic club*. Almqvist Association, Stanford
- Cocconi G, Morrison P (1959) Searching for interstellar communications. *Nature* 184:844–846
- Drake F, Sobel D (1992) *Is anyone out there?* Delacorte Press, New York
- Ewen HI, Purcell EM (1951) Radiation from galactic hydrogen at 1420Mc/s. *Nature* 168:356–357
- Gold T (1968) Rotating neutron stars as the origin of the pulsating radio sources. *Nature* 218:731–732

- Hewish A, Bell SJ, Pilkington JDH, Scott PF, Collins RA (1968) Observation of a rapidly rotating radio source. *Nature* 217:709–713
- Howard A, Horowitz P, Coldwell C, Kleins S, Sng A, Wolff J, Caruso J, Latham D, Papaliolios C, Stefanic R, Zajac J (2000) Optical SETI at Harvard-Smithsonian. *ASP Conf Ser* 213:545–552
- Kardashev N (1964) Transmission of information by extraterrestrial intelligence. *Sov Astr* 8:217
- Lorimer D, Bailes M, McLaughlin MA, Naekevic DJ, Crawford F (2007) A bright millisecond radio burst of extragalactic origin. *Science* 318:777–780
- Paris A, Davies E (2017) [arXiv:1706.04642](https://arxiv.org/abs/1706.04642)
- Ryle M (1962) The new Cambridge radio telescope. *Nature* 194:517–518
- Science: Project Ozma (1960) *Time*. Apr. 18
- Tarter J (2001) The search for extraterrestrial intelligence. *Ann Rev Astron Astrophys* 39:511–548
- Ulmschneider (2003) *Intelligent life in the universe*. Springer, Berlin

Chapter 31

Possible Cultural Impact of Extraterrestrial Life, if It Were to Be Found



Junichi Watanabe

Abstract Detection of extraterrestrial life may soon be within our reach, as shown in the previous chapters of this book. We are at the dawn of the next generation of instruments and space missions aiming to detect evidence of life on other planets. From a social perspective, the situation places astronomy and related fields in a special position among the sciences, due to the public expectations of alien life. This is revealed by a statistical analysis of media exposure which is described in this chapter, along with several expected cultural and social impact should life to be discovered.

Keywords Extraterrestrial life · Press releases · Social and cultural impact

31.1 Introduction

There are two possible cases for the discovery of extraterrestrial life. One is the direct discovery of any life-forms within our solar system, and the other is an indirect discovery of life beyond our solar system. The former may happen during exploration on Mars or by future sample return missions of material ejected by the geysers of the icy satellites that are believed to contain deep oceans under the surface layer of ice, such as Enceladus or Europa. Actual exploration plans are now under serious consideration.

The latter case may be any indirect detection of a so-called biomarker from an exoplanet, especially Earth-like planets within the habitable zone (Kopparapu et al. 2013), by the next-generation large telescopes. There are three plans for building 30–40 m diameter class telescopes, specifically the European Extremely Large Telescope, the Thirty Meter Telescope, and the Giant Magellan Telescope. One of the potential targets of all these telescopes is the search for biomarkers, which are indirect evidence of the existence of extraterrestrial life on an exoplanet. One possible biomarker is the presence of oxygen or ozone in the atmosphere. These

J. Watanabe (✉)
National Astronomical Observatory of Japan, Tokyo, Japan
e-mail: jun.watanabe@nao.ac.jp

molecular species are generally thought to be products of life, although such species are also known to be produced abiotically (Narita et al. 2015). Another biomarker is chlorophyll, which can be detected from the light reflected by the planet's surface. The existence of absorption corresponding to the chlorophyll wavelengths, if it were to be confirmed, is an indirect evidence of extraterrestrial life that resembles plants on the Earth. The effective wavelengths for photosynthesis may be different from that on Earth because the central star in an exoplanet system can have various spectral types. In either case, such a detection will have social and cultural impacts on the general public.

Of course, the direct detection of any artificial radio or laser signals, which is the target of the SETI observations, would have a stronger impact to our culture and society; however, it is beyond the scope of this article.

31.2 Media Exposure of Press Releases Related to Extraterrestrial Life

Astronomy has long inspired the general public, placing the field in a special social position compared with other natural sciences. Press releases related to astronomy are often picked up by many kinds of media. The International Astronomical Union (IAU)'s "International Year of Astronomy" carried out in 2009, the 400-year anniversary of Galileo's first recorded observations with a telescope, caused a huge movement throughout the world and succeeded to include more than 100 countries, which is the top record for UNESCO's commemorative years (Russo et al. 2009).

The image of astronomy in society has been studied by social scientists in Japan. Compared with other sciences, astronomy is regarded as beautiful, easily understandable, and vastness (Toyosawa et al. 2011).

Among astronomy and astrophysics, people are interested in possible extraterrestrial life and exoplanets. This is clearly seen in the levels of media exposure, even when compared with other topics in astronomy. The news of the possible fossils discovered in the Martian Meteorite ALH84001 in 1997 became one of the most debated topics at that time, and controversial discussion still continues on the topic, including the origins of the magnetite nanocrystals (Thomas-Keprta et al. 2009).

Another example was the discovery of the Earth-sized planet around Proxima Centauri announced on August 25, 2016, which was covered by 30 articles in Japanese newspapers. The articles covered not only immediate news release but also short articles encompassing detailed explanations on the background of the present status of this research field. Most of the articles referred to the possible existence of an atmosphere, water, and life on the planet. For comparison, the average number of articles based on press releases from the National Astronomical Observatory of Japan (NAOJ) was just 1.7 per each release during 2016. For example, only 8 articles appeared based on the release dated August 2, 2017, on the discovery of 11 dwarf galaxies and 2 star-containing halos in the outer region of a large spiral galaxy

25 million light-years away from Earth, carried out by Japan's Subaru Telescope. The situation is the same in international media. News on Earth-sized exoplanets has made it into the BBC's yearly science highlights in 2017, 2016, and 2015 (<http://www.bbc.com/news/science-environment-41972289>, <http://www.bbc.com/news/science-environment-38294194>, and <http://www.bbc.com/news/science-environment-35158890>). Exoplanets or astrobiology stories are in the top 10 for *Discover* magazine's top 100 stories for 2017 and 2016 and the top 30 stories in 2015 (<http://discovermagazine.com/2018/janfeb>, <http://discovermagazine.com/2017/janfeb>, <http://discovermagazine.com/2016/janfeb>). Regarding the TRAPPIST-1 system discovery, there was even a Google Doodle (<https://www.google.com/doodles/seven-earth-size-exoplanets-discovered>). These circumstances indicate that the general public is interested in extraterrestrial life and the related topics such as exoplanets.

31.3 Expectation on Change of Concepts

If extraterrestrial life were to be found directly or indirectly, the human concept of life or the world would be changed. Historically, the development of astronomy had led to major shifts in our concept of the world we live. We can find such a case in the Copernican Universe of the seventeenth century. The center of the Universe changed from the Earth to the Sun, namely, from the geocentric to the heliocentric model. The heliocentric model described in Copernicus's book *De revolutionibus orbium coelestium (On the Revolutions of the Celestial Spheres)* in 1543 was not immediately believed but gradually gained credence over a hundred years (Gingerich 2004). After that, it has been revealed that the Sun is not located in the center of the Universe, because the stars in the night sky were shown to be objects similar to the Sun except for the distance from the Earth. In the early twentieth century, the concept of our Galaxy was created, and it has been revealed that the Sun is far out from the central region of the Galaxy. This conceptual change leads us to realize that our Earth is not any special place in this Universe.

Conceptual changes among the general public used to take a long time to manifest after an idea appeared. However, due to the development of the Internet and similar rapid methods of delivering new information, any new concepts or discoveries tend to spread rapidly in the present society. If extraterrestrial life were to be found, we would expect the news to spread swiftly all over the world, and a conceptual change on life or the world would follow. This will take exceedingly short time compared with previous cases such as the heliocentric model. Therefore, we can expect a rapid response to the discovery in various forms to happen. We will finally realize that we are not alone.

This concept change would result in human culture to evolve into a more matured version. As individuals, we become aware of the existence of others during the mental growth from a child to an adult. We ourselves are the center of the world during our infancy. The house where the young infant resides is its whole world. However,

as he or she grows up and leaves the house, he or she learns about the existence of other houses outside, similar to the discovery of other worlds. When the child enters school, he or she is forced to know many other people of the same age and grow further by interacting with their classmates. Finally, he or she will understand that he or she is just one of them and is not at the center of the world. Our cultural evolution may follow the same path. We are the only culture or life we know in our Universe now. If extraterrestrial life were to be found, it would force our culture to mature.

A possible impact to religion would also be expected. For example, the forecast of a bright comet approaching the Earth once made a great impact on members of the cult religion, the Heaven's Gate, of which 39 members died in a mass suicide in Rancho Santa Fe, California, USA, on March 27, 1997. These people believed that comet Hale–Bopp would be followed by a spaceship which their souls could board. Similar extreme cases may occur if extraterrestrial life were to be found. Because of the lack of knowledge regarding the scale of the Universe, it is likely that people would be terrified of attack by the extraterrestrial life. It is definitely important for us to help the general public understand correctly that any evidence of extraterrestrial life is harmless to humankind because of its distance. We need to avoid the tragedy induced by misunderstanding through outreach activities. The first thing we scientists can do is to release timely and accurate information to the media and general public without using exaggerated terms. Exoplanet discoveries are frequently over-hyped by using terms like “second Earth,” “most habitable planets,” and “best candidates for finding life.” Even in the scientific community, “habitable zone” is often used in a misleading way, and we risk losing the public trust by using such careless language (Tasker et al. 2017). At least our science community should be extremely careful using accurate and trustable terms for the news release.

31.4 Conclusion

Extraterrestrial life and its related topics can be the focus of the public and media interest. As we are getting close to be able to detect life on another planet, it is necessary to prepare for the discovery, because the discovery and the subsequent public release will have strong cultural and societal impacts.

References

- Kopparapu RK, Ramirez R, Kasting JF et al (2013) Habitable zones around main-sequence stars: new estimates. *Astrophys J* 765:131. (16pp. <https://doi.org/10.1088/0004-637X/765/2/131>)
- Narita N, Enomoto T, Masaoka S et al (2015) Titania may produce abiotic oxygen atmospheres on habitable exoplanets. *Sci Rep* 5:13977. <https://doi.org/10.1038/srep13977>
- Gingerich O (2004) *The book nobody read*. William Heinemann, London

- Russo P, Cesarsky C, Christensen L (2009) SpS2-The International Year of Astronomy 2009. In: Highlights of astronomy 5(H15):p559–609
- Tasker E, Tan J, Heng K et al (2017) The language of exoplanet ranking metrics needs to change. Nat Astron 1:0042. <https://doi.org/10.1038/s41550-017-0042>
- Thomas-Keppta KL, Clemett SJ, McKay DS et al (2009) Origins of magnetite nanocrystals in Martian meteorite ALH84001. Geochim Cosmochim Acta 73(21):6631–6677
- Toyosawa J, Karasawa K, Todayama K (2011) J Sci Technol Stud 8:151–170 (In Japanese)