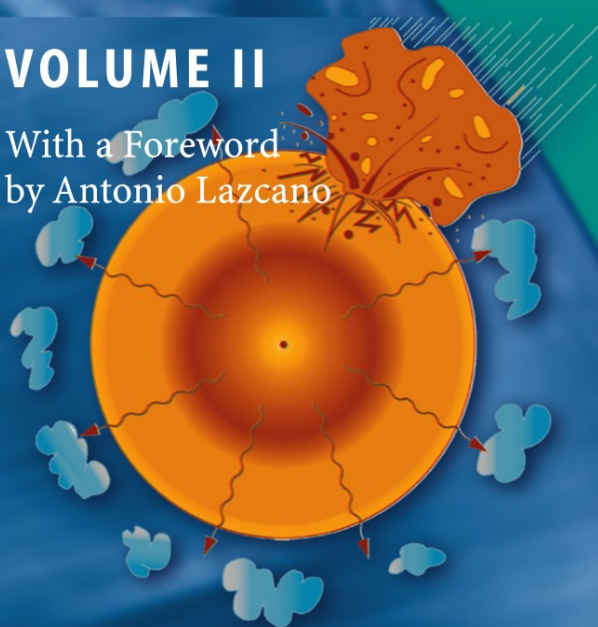


Muriel Gargaud
Hervé Martin · Philippe Claeys
Editors

Lectures in Astrobiology

VOLUME II

With a Foreword
by Antonio Lazcano



 Springer

Lectures in Astrobiology II

Advances in Astrobiology and Biogeophysics

springer.com

This series aims to report new developments in research and teaching in the interdisciplinary fields of astrobiology and biogeophysics. This encompasses all aspects of research into the origins of life – from the creation of matter to the emergence of complex life forms – and the study of both structure and evolution of planetary ecosystems under a given set of astro- and geophysical parameters. The methods considered can be of theoretical, computational, experimental and observational nature. Preference will be given to proposals where the manuscript puts particular emphasis on the overall readability in view of the broad spectrum of scientific backgrounds involved in astrobiology and biogeophysics.

The type of material considered for publication includes:

- Topical monographs
- Lectures on a new field, or presenting a new angle on a classical field
- Suitably edited research reports
- Compilations of selected papers from meetings that are devoted to specific topics

The timeliness of a manuscript is more important than its form which may be unfinished or tentative. Publication in this new series is thus intended as a service to the international scientific community in that the publisher, Springer, offers global promotion and distribution of documents which otherwise have a restricted readership. Once published and copyrighted, they can be documented in the scientific literature.

Series Editors:

Dr. André Brack
Centre de Biophysique Moléculaire
CNRS, Rue Charles Sadron
45071 Orléans, Cedex 2, France
Brack@cnsr-orleans.fr

Dr. Gerda Horneck
DLR, FF-ME, Radiation Biology
Linder Höhe
51147 Köln, Germany
Gerda.Horneck@dlr.de

Prof. Dr. Michel Mayor
Observatoire de Genève
1290 Sauverny, Switzerland
Michel.Mayor@obs.unige.ch

Dr. Christopher P. McKay
NASA Ames Research Center
Moffet Field, CA 94035, USA

Prof. Dr. H. Stan-Lotter
Institut für Genetik
und Allgemeine Biologie
Universität Salzburg
Hellbrunnerstr. 34
5020 Salzburg, Austria

Muriel Gargaud

Hervé Martin Philippe Claeys (Eds.)

Lectures in Astrobiology

Volume II

With a Foreword by Antonio Lazcano

and 42 Tables and 215 Figures Including Photos and Plates, 44 in Color

Dr. Muriel Gargaud
Editor-in-Chief
Université Bordeaux 1
Observatoire Aquitain
des Sciences de l'Univers
2 Rue de l'Observatoire
33270 Floirac, France

Prof. Hervé Martin
Université Blaise Pascal
Laboratoire Magmas et Volcans
5 Rue Kessler
63038 Clermont-Ferrand, France

Prof. Philippe Claeys
Vrije Universiteit Brussel
Department of Geology
Pleinlaan 2
1050 Brussels, Belgium

Muriel Gargaud et al. (eds.), *Lectures in Astrobiology II, Adv. Astrobiol. Biogeophys.* (Springer, Berlin Heidelberg 2007), DOI 10.1007/10913314

Library of Congress Control Number: 2005937604

ISSN 1610-8957

ISBN-10 3-540-33692-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-33692-1 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media
springer.com

© Springer-Verlag Berlin Heidelberg 2007

The use of general descriptive names, registered names, trademarks, 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.

Typesetting and production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover design: Erich Kirchner, Heidelberg

Printed on acid-free paper

55/3100/YL - 5 4 3 2 1 0

« Voyez-vous cet œuf? C'est avec cela qu'on renverse
toutes les écoles de théologie et tous les temples de la terre. »

Denis Diderot (*Entretien avec d'Alembert*)

Foreword

Long before the idea of spontaneous generation was incorporated by Jean-Baptiste de Lamarck into evolutionary biology to explain the first emergence of life, the possibility that other planets were inhabited had been discussed, sometimes in considerable detail, by scientists and philosophers alike. More often than not, these were speculations that rested on the idea of a uniform universe but with little or no empirical basis. Today our approaches to the issue of life in the Universe have changed dramatically; neither the formation of planets nor the origin of life are seen as the result of inscrutable random events, but rather as natural outcomes of evolutionary events. The interconnection between these two processes is evident: understanding the formation of planets has major implications for our understanding of the early terrestrial environment, and therefore for the origin of living systems.

Although it is tempting to assume that the emergence of life is an unavoidable process that may be continuously taking place throughout the Universe, it is still to be shown that it exists (or has existed) in places other than the Earth. With the exception of Mars and some speculations on Europa, prospects for life in our own solar system have been strongly diminished. Although there is evidence the early Martian environment was milder and may have been similar to the primitive Earth, today its surface is a deep-frozen desert, constantly bathed by short-wavelength ultraviolet radiation. This highly oxidizing environment has rendered any hypothetical biosphere extinct or has limited it to few restricted underground niches well below the surface, where brine aquifers appear to be present. I am one of those sadly convinced that the balance of evidence suggests that life in our planetary system is confined to our own planet. As shown by the debates sparked by the announcement that the Allan Hills 84001 meteorite included traces of ancient Martian life, we also lack a well-defined consensus regarding the criteria by which we could rapidly recognize evidence of extraterrestrial biological activity.

Recognition that meteorite impacts may have led to an intense exchange of rocky ejecta between the inner planets during the early phases of the solar system has led some to discuss the possibility that life on our planet may have an ultimate Martian origin. It is somewhat amusing to see that discussions on panspermia, i.e., the transfer of organisms from one planet to another, are periodically resurrected without providing any detailed explanations of the

ultimate mechanisms which may have led to the appearance of life in extraterrestrial habitable environments. It is true that the high UV-resistance of different prokaryotic species at the low temperatures of deep space, the likelihood of artificial or directed transport of microorganisms by probes sent to other bodies in the solar system, and the recognition of the Martian origin of some meteorites have given additional support to the panspermia hypothesis. However, this only shifts the problem to a different location, and most researchers prefer to study the origin of life within the historical framework of an evolutionary analysis that assumes that it took place on Earth.

As shown by the chapters that form this volume, nowadays the genealogy of life can be extended back to the origin of the chemical elements, continues with the evolution of stars and the formation of planets, and continues further with the synthesis of organic compounds that are found in comets and meteorites, which show that during the time of formation of the Earth and other planets the synthesis of many organic compounds which we associate today with living systems was taking place. Although we do not know how the transition from the non-living to the living took place, today the phylogenetic analysis of genomes can provide us with a historical record that very likely can be extended prior to the divergence of the three extant cell lineages. Most of the modern scenarios start out with relative simple organic molecules, now known to be widely distributed, which are readily synthesized, and hypothesized to undergo further evolutionary changes leading into self-maintaining, self-replicating systems from which the current DNA/protein-based biology resulted. Although many open questions remain, it is reasonable to conclude that life is the natural outcome of an evolutionary process, and that it may have appeared elsewhere in the Universe.

The distinguished American evolutionist George Gaylord Simpson once wrote that “exobiology is still a science without any data, therefore no science.” Can the same be said today of astrobiology? The idea that life is the result of a rare chance event has been replaced by an evolutionary narrative, according to which biological systems are the outcome of a gradual but not necessarily slow process that began with the abiotic synthesis of biochemical monomers and eventually led to self-sustaining, self-replicating systems capable of undergoing Darwinian evolution. There is no compelling reason to assume that such processes took place only on the Earth. The timescale for the origin and early evolution of life and the ease of formation of amino acids, purines, and other biochemical compounds under a relatively wide range of reducing conditions, together with the abundance of organic molecules throughout space, all speak for natural laws conducive to the emergence of life in extraterrestrial environments where similar conditions prevail.

Yet, the role of historical contingency cannot be discounted. As the French philosopher Pascal once remarked, had Cleopatra’s nose been different, the course of history may have changed. Precellular evolution was not a continuous, unbroken chain of progressive transformations steadily proceeding to the

first living systems. Many prebiotic cul-de-sacs and false starts probably took place. While it may be true that the transition to life from non-living systems did not require a rather narrow set of environmental constraints, we cannot discount the possibility that even a slight modification of the primitive environment could have prevented the appearance of life on our planet. However unpalatable this conclusion may be, life may be a rare and even unique phenomenon in the Universe. In fact, today we have no evidence of extraterrestrial life, and we should not forget that it is like democracy: everybody likes the idea and speaks about it, but no one has really seen it.

Mexico City, February 2006

Antonio Lazcano

Preface

This is the second book dedicated to the origin(s) and development of life on Earth and possibly elsewhere in the Universe. It continues and supplements *Lecture in Astrobiology, vol. 1* published in 2005. The main goal of these volumes is to present the current state of knowledge concerning the environmental conditions and the processes leading to the appearance of life, and to establish the parameters indicative of biological activity on ancient Earth and eventually on other planetary bodies.

This book summarizes the lectures presented by selected speakers during Exobio'03, Ecole d'exobiologie du CNRS held in September 2003 in Propriano, Corsica. Just as in this volume, the field of exobiology is by nature multidisciplinary. It discusses the *bio-geo-physico-chemical* conditions required for the origin, development and evolution of life on planet Earth. Consequently, it addresses also the possibility that forms of life may exist (or have existed, or will exist) elsewhere in the solar system or in the rest of the Universe. These themes are often referred to during the in situ exploration of Mars and Titan or the ongoing search for distant exoplanets.

Recent geological investigations and in particular the discovery of the 4.4–4.3 Ga old Jack Hills zircons in Australia, demonstrate that part of the conditions required for the emergence of life existed on Earth shortly after the end of the accretion period. The first chapters deal with the different processes responsible for the establishment of the primitive environments on the young Earth. The stellar genesis and the distribution mechanisms of the key chemical elements (C, O, Si, Ca, Fe) available for the formation of the planets are discussed. An attempt is made to unravel the precise chronology of the astronomical and geological processes, which from the planetary accretion to the development of the crusts have paved the way to an environment propitious for life. Isotopic data obtained from various mineral phases in meteorites as well as the frequency and intensity of asteroid and comet impacts constrain these events.

So far, life seems directly linked to the presence of liquid water. However, it is now demonstrated that it can flourish in a wide variety of terrestrial environments, some at first sight highly inhospitable. The fact that liquid water most likely existed at the surface of Mars several billion years ago raises the challenging and so far unanswered question of the birth and development of life on this planet. Therefore, it also triggers discussions about its preservation

and complete extinction due the geological evolutionary path followed by Mars. An analog might be provided by mass extinction events on Earth. The ongoing studies of Titan, which its atmosphere and surface could represent, according to some authors, a “laboratory of prebiotic chemistry”, demonstrate the diversity and high complexity of the available planetary conditions. The cases of Mars and Titan directly address the concept of habitability; a notion that appears to be different (but somehow complementary) for the astronomer and the biologist. However, a favorable environment alone is not sufficient for life; the essential chemical building blocks must also be present. This particular aspect is considered in two chapters focused on the modeling of the type of molecules existing in interstellar clouds and planetary atmospheres, and on the regulation of life by planetary setting. In a completely different approach, clearly indicative of the inner intricacy of the essence of life, its artificial form is discussed, as it represents the ultimate concept of habitability and evolution inside a computer program.

This book was written for a large public of scientists as well as students, interested in the different challenges presented by the origin of life, its development and its possible existence outside the realm of Earth. It includes several appendices and an extensive glossary, to complement or update the reader’s knowledge in the many disciplines covered. The different chapters are condensed versions of the animated discussions held in Propriano by a community of astronomers, geologists, chemists, biologists and computer scientists, all sharing the common goal to establish and evaluate potential scenarios leading to the appearance and development of life. This book attempts to convey the enthusiasm, the vigor, and the richness of the debates generated when specialists from a variety of disciplines gather their strength to address a specific and challenging theme. Dogmas break apart, and new paths are explored as everyone may come to question some basic principles of their own speciality and must integrate in his/her thinking, knowledge and principles gained from several other disciplines. The astronomer must then learn to reason as a biologist, and the chemist must assimilate geological parameters; the ambition of this book is to promote this broad scientific trespassing.

The editors wish to thank every author, who in his own way contributed a piece of knowledge to what remains an inextricable puzzle, whose complexity increases with every new discovery in one field or the other. The work and the patience of the reviewers is acknowledged; their contributions greatly improved the manuscripts.

*Muriel Gargaud
Philippe Claey's
Hervé Martin*



From *left to right*: Hervé Martin (geochemist), Muriel Gargaud (astrophysicist), Philippe Claeys (geologist)

Contents

1 Stellar Nucleosynthesis

<i>Nikos Prantzos</i>	1
1.1 Introduction	1
1.2 Nuclei in the Cosmos	2
1.2.1 Solar and Cosmic Abundances	2
1.2.2 Cosmic Abundances vs. Nuclear Properties	3
1.2.3 Overview of Nucleosynthesis	6
1.3 Stars: from the Main Sequence to Red Giants	8
1.3.1 Basic Stellar Properties	8
1.3.2 H-Burning on the Main Sequence	10
1.3.3 He-Burning in Red Giants	12
1.4 Advanced Evolution of Massive Stars	15
1.4.1 Neutrino Losses Accelerate Stellar Evolution	15
1.4.2 C, Ne, and O-Burning	17
1.4.3 Si-Melting and Nuclear Statistical Equilibrium (NSE)	21
1.4.4 Overview of the Advanced Evolutionary Phases	24
1.5 Explosive Nucleosynthesis in Supernovae	26
1.5.1 Main Properties and Classification of Supernovae	26
1.5.2 Explosive Nucleosynthesis in Core Collapse Supernovae	27
1.5.3 Explosive Nucleosynthesis in Thermonuclear SN	32
1.5.4 Production of Intermediate Mass Nuclei (from C to the Fe peak)	34
1.6 Nuclei Heavier than Fe	35
1.6.1 Production Mechanisms and Classification of Isotopes	35
1.6.2 The S-Process	37
1.6.3 The R-Process	38
1.7 Summary	40
References	42

2 Formation of the Solar System: a Chronology Based on Meteorites

<i>Marc Chaussidon</i>	45
2.1 In Search for Ages	45
2.2 What is a Geochemical Age?	46
2.2.1 The Radioactivity	46

2.2.2	The Absolute Ages	47
2.2.3	The Relative Ages	48
2.2.4	Sources of Error or Uncertainty in Isotopic Dating	49
2.3	What are the Processes that can be Dated by Isotopic Analyses of Meteorites?	51
2.4	From the First Solids to the First Planets: When and How Fast?....	53
2.4.1	The Age of Meteorites and the Duration of Accretion Processes: the First Order Picture	53
2.4.2	A Relative Chronology Based on the Extinct Radioactivity of ^{26}Al	53
2.4.3	A Relative Chronology Based on the Extinct Radioactivity of ^{53}Mn	57
2.4.4	Absolute Calibration of the ^{26}Al and ^{53}Mn Chronologies.....	58
2.4.5	Longer Period Extinct Radioactivities and Chronology of the Differentiation	60
2.5	Remaining Questions	63
2.5.1	Disparities Between the Various Chronologies ^{26}Al , ^{53}Mn and ^{182}Hf	63
2.5.2	The Hypothesis of Homogeneity of the Distribution of Extinct Radioactivities in the Solar Accretion Disk: the Origin of Extinct Radioactivities	65
2.6	Conclusions	67
	References	69

3 The Formation of Crust and Mantle of the Rocky Planets and the Mineral Environment of the Origin of Life

<i>Francis Albarède</i>	75	
3.1	Chemical and Mineralogical Structure of the Earth.....	75
3.2	Dynamics of the Earth's Interior	77
3.3	The Origin of Continents	80
3.4	The Early Ages of Our Planet	83
3.5	From One Planet to the Next	88
3.6	Some Speculations	98
3.7	Questions for the Future	99
	References	100

4 Water and Climates on Mars

<i>François Forget</i>	103	
4.1	Introduction	103
4.2	Mars' Present-Day Climate	103
4.2.1	The CO_2 Cycle and the Seasonal Polar Caps	105
4.2.2	The Dust Cycle	106
4.2.3	The Water Cycle	107
4.3	A Few Million Years Ago: the Recent Martian Paleoclimates	110
4.3.1	Climate Changes Due to Orbital Parameter Variations	110

4.3.2	Liquid Water on Mars a Few Million Years Ago	113
4.4	More Than Three Billion Years Ago: the Youth of Mars	115
4.4.1	Evidence for Sustained Liquid Water on Early Mars	115
4.4.2	The Early Mars Climate Enigma	118
4.5	Conclusion	119
References		119

5 Planetary Atmospheres: From Solar System Planets to Exoplanets

<i>Thérèse Encrenaz</i>		123
5.1	What is an Atmosphere?	123
5.1.1	Atmospheric Structure	123
5.1.2	Atmospheric Circulation and Cloud Structure	126
5.1.3	Atmospheric Composition	128
5.1.4	Interaction with the Magnetosphere	130
5.2	Atmospheres of Solar System Planets	130
5.2.1	Formation and Evolution of Planetary Atmospheres in the Solar System	130
5.2.2	Terrestrial Planets and Giant Planets	132
5.2.3	Atmospheres of Outer Satellites and Pluto	139
5.3	Tools for Studying Planetary Atmospheres	141
5.3.1	Remote Sensing Analysis	141
5.3.2	<i>In Situ</i> Analysis: Chemical Composition from Mass Spectrometry	146
5.4	From Solar System Planets to Exoplanets	147
5.4.1	Properties of Detected Exoplanets: a Summary	147
5.4.2	Earth-Like Exoplanets: the Habitability Zone	148
5.4.3	Giant Exoplanets: Structure and Composition	149
5.5	Conclusions	151
References		152

6 What About Exoplanets?

<i>Marc Ollivier</i>		157
6.1	Let's Talk About History	157
6.2	Statistical Analysis of the First Extrasolar Planets Discoveries	159
6.2.1	The Mass Distribution of Exoplanets	160
6.2.2	The Star Planet Distance Distribution	162
6.2.3	Orbit Migration	163
6.2.4	Mass/Distance Relation for Exoplanets	171
6.2.5	Eccentricity of Exoplanet Orbits	172
6.2.6	The Metallicity of Stars with Planets	174
6.3	The Atmospheres and Spectra of Giant Exoplanets	175
6.3.1	General Considerations	175
6.3.2	Pegasides: the Point of View of Theoreticians and Observers	176

6.4	Future Steps in Exoplanetology and Associated Instrumentation	180
6.4.1	Open Questions	180
6.4.2	Research and Study of Giant Planets	181
6.4.3	Research and Study of Telluric Planets	183
6.4.4	Characterization of Exoplanets by Direct Detection	186
	References	193

7 Habitability: the Point of View of an Astronomer

	<i>Franck Selsis</i>	199
7.1	Introduction	199
7.2	The Circumstellar Habitable Zone	201
7.2.1	The Inner Limit of the Habitable Zone	202
7.2.2	The Outer Limit of the HZ (or How to Warm Early Mars?) . .	206
7.2.3	Continuously Habitable Zone	210
7.3	Habitability Around Other Stars	210
7.4	The Influence of the Giant Planets on the Habitability of the Terrestrial Planets	214
7.5	Discussion	215
7.6	Conclusion and Perspectives	216
	References	217

8 Habitability: the Point of View of a Biologist

	<i>Purificación López-García</i>	221
8.1	Introduction	221
8.1.1	The Concept of Habitability	221
8.1.2	Habitability in Biological Terms	222
8.2	What is Life?	222
8.2.1	Life's Definitions	222
8.2.2	Is it Living?	223
8.3	The Cell	224
8.3.1	Properties	224
8.3.2	Prokaryotes and Eukaryotes	225
8.3.3	The Tree of Life	226
8.4	Common Denominators of Life on Earth	227
8.4.1	Elements and Molecules	228
8.4.2	Cellular Metabolism	230
8.4.3	The Limits of Terrestrial Life	233
8.5	Perspectives	235
	References	235

9 Impact Events and the Evolution of the Earth

	<i>Philippe Claeys</i>	239
9.1	Introduction	239
9.1.1	Terrestrial Craters	239
9.1.2	Historical Perspective of the Impact Process	241

9.2	Characteristics of Impact Craters	242
9.2.1	Magnitude and Frequency	242
9.2.2	Crater Morphologies	244
9.2.3	Formation Mechanism (Based Essentially on Melosh 1989 and French 1998)	246
9.2.4	Identification Criteria	249
9.3	Case Study: The Cretaceous-Tertiary Boundary and the Chicxulub Crater	254
9.3.1	The Chicxulub Crater	254
9.3.2	Distribution of Ejecta	257
9.3.3	Consequences for the Biosphere	262
9.3.4	Asteroid or Comet?	264
9.4	Stratigraphic Distribution of Impact Events	264
9.4.1	In the Phanerozoic (0 to 540 Ma)	264
9.4.2	Proterozoic Impacts (540 Ma to 2.5 Ga)	266
9.4.3	Archean Impacts (2.5 to 4 Ga)	267
9.4.4	Hadean Impacts (4.0 Ga to Formation of the Earth)	268
9.5	Discussion: Impact, Origin of Life and Extinctions	270
	References	273

10 Towards a Global Earth Regulation

	<i>Philippe Bertrand</i>	281
10.1	The Oxygen: an Energy Story	282
10.2	Nitrogen and Phosphorus: the Nutrient Feedback	287
10.3	What About the Atmospheric CO ₂ ?	291
10.4	Towards a Global Biogeochemical Regulation (Homeostasy)	296
	References	302

11 The Last Common Ancestor of Modern Cells

	<i>David Moreira, Purificación López-García</i>	305
11.1	The Last Common Ancestor, the Cenancestor, LUCA: What's in a Name?	305
11.1.1	Some Historical Grounds	305
11.1.2	The Hypothesis of a Cenancestor	306
11.2	How Did the Cenancestor Make Proteins?	308
11.3	What Was the Nature of the Genetic Material?	309
11.4	What Did the Cellular Metabolism Look Like?	310
11.5	Was the Cenancestor Membrane-Bounded?	311
11.6	Other Unresolved Questions	312
11.7	Perspectives	314
	References	315

12 An Extreme Environment on Earth: Deep-Sea Hydrothermal Vents Lessons for Exploration of Mars and Europa

<i>Daniel Prieur</i>	319
12.1 Some Features of Oceanic Environment	319
12.2 Deep-Sea Hydrothermal Vents	320
12.3 Highly Efficient Symbioses	324
12.3.1 Vestimentifera	325
12.3.2 Molluscs	325
12.3.3 Polychaetous Annelids	327
12.3.4 Crustaceans	328
12.4 Life at High Temperature	328
12.4.1 Novel Microorganisms in the Bacteria Domain	329
12.4.2 Novel Microorganisms in the Archaea Domain	329
12.5 Response to Hydrostatic Pressure	335
12.6 Other Specific Adaptations	336
12.6.1 Fluctuations of Environmental Conditions	336
12.6.2 Heavy Metals	337
12.6.3 Ionizing Radiations	337
12.7 Lessons from Microbiology of Hydrothermal Vents	337
References	342

13 Comets, Titan and Mars: Astrobiology and Space Projects

<i>Yves Bénilan, Hervé Cottin</i>	347
13.1 An Astrobiological Look at the Solar System	348
13.1.1 The Origin of the Organic Matter	348
13.1.2 Follow the Water	354
13.2 The Space Exploration of Comets	356
13.2.1 General Considerations	356
13.2.2 Past Missions	359
13.2.3 Current Missions	369
13.2.4 Future Space Missions	382
13.3 The Space Exploration of Titan	382
13.3.1 Observations and Models of Titan Before Space Missions	382
13.3.2 Voyager Missions at Titan	385
13.3.3 Similarities and Differences Between Titan and the Earth	388
13.3.4 Cassini–Huygens Mission	392
13.4 Mars Exploration	398
13.4.1 Mars Before Space Missions	398
13.4.2 The Beginning of Martian Exploration	399
13.4.3 Current Space Missions	408
13.4.4 Future Exobiological Missions	412
13.5 Conclusion	420
References	420

14 Quantum Astrochemistry:

Numerical Simulation as an Alternative to Experiments

<i>Yves Ellinger, Françoise Pauzat</i>	429
14.1 The Methods of Quantum Chemistry	429
14.1.1 Definition of Quantum Chemistry Calculations and Approximations	429
14.1.2 Wave Function-Based Methods	437
14.1.3 DFT Methods	447
14.1.4 Wave Function Versus DFT Methods	454
14.2 Applications of Quantum Chemistry	455
14.2.1 Radio Millimeter Observations	458
14.2.2 Infrared Observations	463
14.2.3 Modeling of Chemical Processes	472
14.2.4 Exobiology	482
14.3 Conclusions and Prospective	485
References	487

15 Artificial Life or Digital Dissection

<i>Hugues Bersini</i>	491
15.1 Introduction to Artificial Life	491
15.2 The History of Life Seen by Artificial Life	500
15.2.1 Appearance of a Chemical Reaction Looped Network	500
15.2.2 Production by this Network of a Membrane Promoting Individuation and Catalyzing Constitutive Reactions	502
15.2.3 Self-Replication of the Elementary Cell	503
15.2.4 Genetic Coding and Evolution by Mutation, Recombination and Selection	506
15.3 Functional Emergence	509
15.3.1 Emergence Within Networks: a Short Introduction to the Three Networks Studied at IRIDIA	512
15.3.2 Small Worlds	525
15.3.3 Emergence in Cellular Automata	529
15.3.4 Useful Emergence	534
15.4 Plasticity and Adaptability	535
15.5 Environmental Autonomy and Significant Integration	540
15.6 Conclusions	542
References	544

Appendix

1 Some Astrophysical Reminders

<i>Marc Ollivier</i>	549
1.1 A Physics and Astrophysics Overview	549
1.1.1 Star or Planet?	549

1.1.2	Gravitation and Kepler's Laws	550
1.1.3	The Solar System	550
1.1.4	Black Body Emission, Planck Law, Stefan–Boltzmann Law . .	551
1.1.5	Hertzsprung–Russel Diagram, the Spectral Classification of Stars	553
1.2	Exoplanet Detection and Characterization	555
1.2.1	Planet Detection by the Radial Velocity Method	555
1.2.2	Planet Detection by Astrometry	557
1.2.3	Planet Detection by the Transit Method	558
1.2.4	Exoplanet Direct Detection by Nulling Interferometry (Bracewell's Interferometer)	560
1.3	List of Exoplanets as Detected the 31th of January 2005 (Schneider 2004; see References in Chap. 6)	562
2	Useful Astrobiological Data	567
2.1	Physical and Chemical Data	567
2.2	Astrophysical Data	574
2.3	Geological Data	579
2.4	Biochemical Data	589
3	Glossary	595
	Authors	659
	Index	665

1 Stellar Nucleosynthesis

Nikos Prantzos

1.1 Introduction

The theory of nucleosynthesis emerged around the middle of the twentieth century as a result of rapid progress in our understanding of three different fields:

1. The composition of the Sun and the solar system
2. The physical conditions prevailing in the interiors of stars during their various evolutionary stages
3. The systematic properties of nuclei and nuclear reactions

The idea that all nuclei are synthesized in the hot stellar interiors was promoted by the British astronomer Fred Hoyle in the 1940s. On the other hand, at about the same period, the Russian physicist George Gamow argued that all nuclei were produced in the hot primordial universe¹, by successive neutron captures. However, it was rapidly shown that in those conditions it was extremely difficult to synthesize anything heavier than ^4He . Moreover, observations in the 1950s showed that although all stars have similar amounts of the light (and most abundant) elements H and He, they may differ considerably in their heavy element content. It was clear then that heavy nuclei had to be produced *after* the Big Bang, in successive stellar generations, which progressively enriched the galaxies.

In the following, a brief account is presented on the basic ideas underlying nucleosynthesis. Emphasis is given (Sects. 1.4 and 1.5) on nucleosynthesis in massive stars, which synthesize most of the abundant heavy elements (C, O, Si, Ca, Fe, etc.) that are important for the formation of terrestrial planets and for the emergence of life.

¹ The name *Big Bang* was ironically given to the theory of the hot early universe by Hoyle, who promoted instead the *steady state theory* for the universe. Hoyle was proven to be wrong in his cosmological views, but correct as to the origin of the elements; the opposite happened with Gamow.

1.2 Nuclei in the Cosmos

1.2.1 Solar and Cosmic Abundances

According to our current understanding, the material of the protosolar nebula had a remarkably homogeneous composition, as a result of high temperatures (which caused the melting of the quasitotality of dust grains) and thorough mixing. This composition characterizes the present-day surface layers of the Sun, which remain unaffected by nuclear reactions occurring in the solar interior (with a few exceptions, concerning, e.g., the fragile D and Li). The abundances of most elements in the solar photosphere are now established to a fairly good precision.

Once various physicochemical effects are taken into account² it appears that the elemental composition of the Earth and meteorites matches extremely well the solar photospheric composition.³ On the other hand, Earth and meteoritic materials provide the opportunity of measuring their isotopic composition with extreme accuracy in the laboratory, while such measurements are, in general, impossible in the case of the Sun.⁴

A combination of solar and meteoritic measurements allows us to establish the solar composition (Fig. 1.1), presumably reflecting the one of the protosolar nebula 4.5 Gyr ago. An early attempt to obtain such a curve was made by Goldsmith in 1938, while Suess and Urey provided in 1956 the first relatively complete and precise data set, on which the founding works of nucleosynthesis (Burbidge et al. 1957, Cameron 1957) were based. The most recent major compilations are the ones of Lodders (2003) and Asplund et al. (2004).

In the early 1950s it was realized that the composition of stars in the Milky Way presents both striking similarities and considerable differences with the solar composition. The universal predominance of H (90% by number) and He (9% by number) and the relative abundances of “metals” (elements heavier than He, with O, C, N, Ne, Fe being systematically more abundant than the other species) is the most important similarity. On the other hand, the fraction of metals (metallicity) appears to vary considerably (Fig. 1.2), either within the solar vicinity (where the oldest stars have a metallicity of 0.1 solar), across the Milky Way disk (with stars in the inner galaxy having 3 times more metals than

² For instance, the low gravity of the Earth and smaller bodies of the solar system was insufficient to retain the light H and He of the protosolar mixture; despite their high initial abundances, these elements are absent from those bodies (the H combined to chemically reactive O in the form of water was added to the Earth after its formation, during a later accretion period).

³ Some long standing discrepancies between meteoritic and solar measurements concerning the abundances of several key elements were solved in the 1990s (for Fe) and in the early 2000s (for O and C).

⁴ Nuclear spectroscopy, through the detection of characteristic γ -ray lines from the de-excitation of nuclei in solar flares, offers a (limited) possibility of determining the Sun’s isotopic composition.

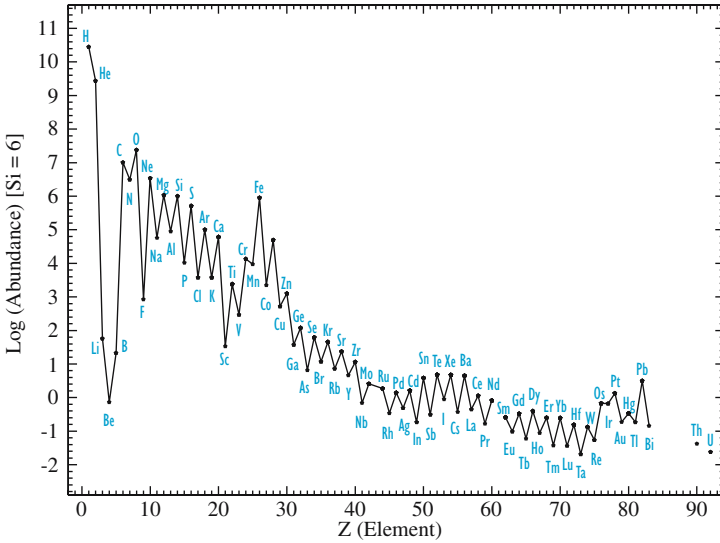


Fig. 1.1. Solar system abundances by number, in a scale where $\log(S_i) = 6$, as a function of the charge number Z of the element. H and He are by far the most abundant elements (90% and 9%, respectively, by number, or 70% and 28% by mass). Note also the extremely low abundances of Li, Be and B with respect to the neighboring C, N and O; the peak in the abundance curve around Fe; the “saw-tooth” overall pattern; and the high abundance of Pb

the Sun) or in the galactic halo (with stellar metallicities ranging from 0.1 to 0.00001 solar).

These variations in composition are extremely important for understanding the “chemical evolution” of the Milky Way. Indeed, they reflect the progressive enrichment of the various components of the galaxy (halo, bulge, disk) with metals produced and ejected by successive stellar generations. The first generation was presumably formed from gas of primordial composition, i.e., H and He (with a trace amount of Li) resulting from the early hot universe of the Big Bang. However, the large diversity of the metallicity of stars (also observed in the interstellar medium of the Milky Way and other galaxies) should not mask the important fact of uniformity in the basic pattern, namely the predominance of H and He and the quasiuniformity (within a factor of a few) in the abundance ratios between metals. It is precisely that uniformity which requires an explanation, involving nuclear reactions in appropriate astrophysical sites.

1.2.2 Cosmic Abundances vs. Nuclear Properties

The solar or cosmic abundances of the various nuclear species (Fig. 1.3, upper panel) constitute an important macroscopic property of (baryonic) matter. It

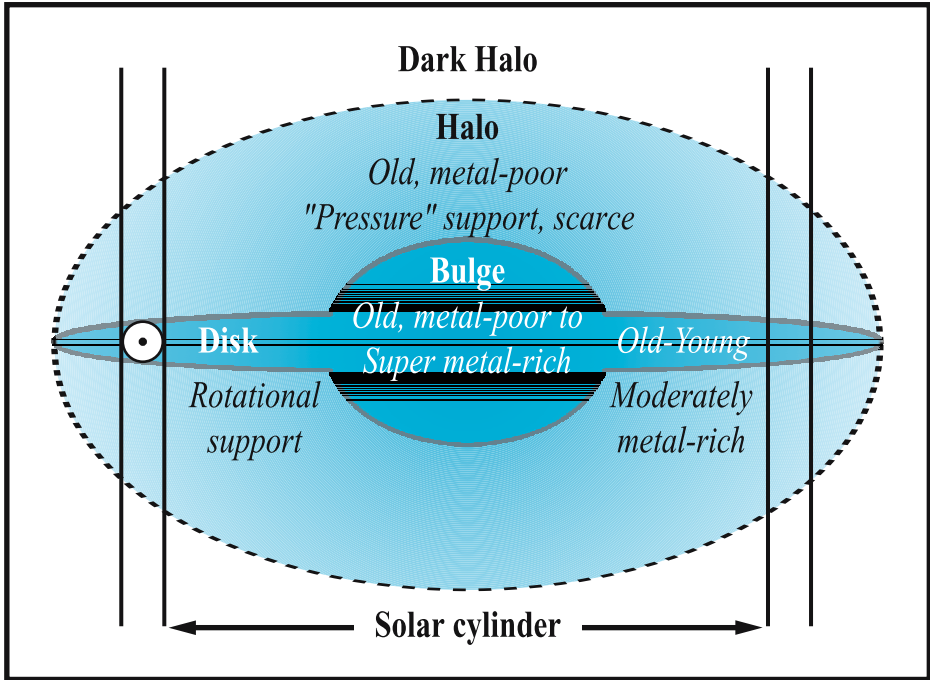


Fig. 1.2. Schematic cross-section of the Milky Way, with the Sun at a distance of 25 000 light-years from the galactic center and the various galactic components (halo, bulge, disk) displayed, along with the main properties of the corresponding stellar populations

was realized that this property is closely related to a microscopic one, namely the binding energy per nucleon (BEN) (see lower panel in Fig. 1.3). This quantity represents the energy per nucleon required to break a nucleus in its constituent particles, and is a measure of the nuclear stability. In the framework of the liquid drop model of the nucleus (composed of A nucleons), it is described approximately by the formula of Bethe and Weizsaecker (1935):

$$B(A, Z) = f_V - f_S A^{-1/3} - f_E Z(Z - 1) A^{-4/3} - f_{SM} (A - 2Z)^2 / A^2 \quad (1.1)$$

According to that formula, the BEN results from the competition (or synergy) of several factors:

- The short range nuclear attraction between neighboring nucleons, contributing a constant term f_V (where V stands for volume, the corresponding contribution to the total binding energy of the nucleus $BE = BEN \times A$ being proportional to the number of nucleons A , which occupy a volume V).
- A symmetry term (with coefficient f_{SM}), arising in part from the Pauli exclusion principle and in part from symmetry effects in the nucleon–nucleon

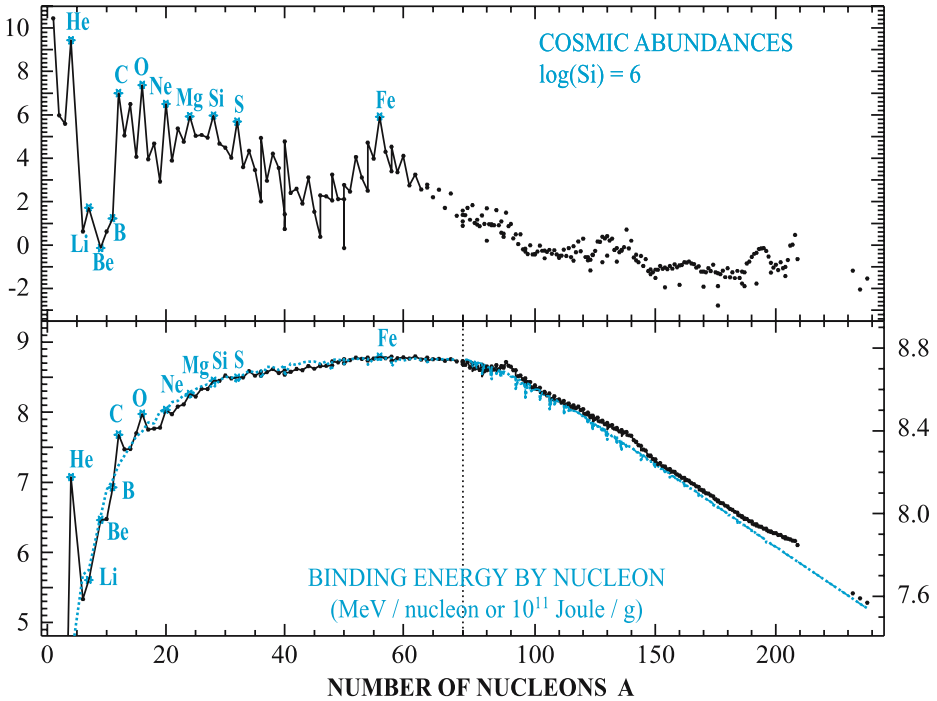


Fig. 1.3. Cosmic abundances (*top panel*) and nuclear binding energy per nucleon (*bottom panel*) as a function of nuclear mass number A . Strongly bound nuclei (“ α nuclei”, like ${}^4\text{He}$, ${}^{12}\text{C}$, ${}^{16}\text{O}$, ${}^{20}\text{Ne}$, ${}^{24}\text{Mg}$, ${}^{28}\text{Si}$, or Fe-peak nuclei) are more abundant than their neighbors. In the *lower panel*, the *continuous curve* connects experimental data points, while the *dotted curve* is a straight application of the liquid drop model (see text), without quantum-mechanical corrections. Note the change in the horizontal scale at $A = 70$, as well as the different vertical scales in the right and left parts of the lower panel

interaction, which favors equal numbers of protons and neutrons, as well as even rather than odd nuclei.

- The long range electrostatic repulsion between protons (with coefficient f_E), which favors an increasing fraction of neutrons when the mass number A increases.
- A surface term (with coefficient f_S), representing a reduced contribution in the binding energy from nucleons at the “surface” of the nucleus. The importance of this term decreases with increasing A (i.e., with decreasing surface to volume ratio).

A simple calculation of the BEN along those lines reproduces satisfactorily the gross features of the measured curve (see Fig. 1.3): BEN increases steadily up to the Fe peak (due to decreasingly important negative contributions from the surface term) and then declines slowly (due to increas-

ingly important negative contributions from the electrostatic and symmetry terms).

However, reproducing several key features of the BEN curve requires a full quantum-mechanical treatment. This is done in the framework of the nuclear shell model, which assumes that each nucleon is moving in the potential created by all the other nucleons. That treatment leads to quantized energy levels (Fig. 1.4) and accounts for the propensity of identical nucleons to form pairs with opposite spins, which introduces a supplementary (positive) contribution to the BEN curve; it also accounts for the exceptional stability of α nuclei (with nucleon number in multiples of 4, like ${}^4\text{He}$, ${}^{12}\text{C}$, ${}^{16}\text{O}$, ${}^{20}\text{Ne}$, ${}^{24}\text{Mg}$, etc.) and for the stability of nuclei with “magic” nucleon numbers 2, 8, 20, 28, 82, 126 (corresponding to filled nuclear shells in Fig. 1.4).

The key features of the BEN curve are obviously reflected in the cosmic abundance curve, albeit at a local level only: more stable nuclei are more abundant than their neighbors (e.g., the α nuclei or the Fe-peak nuclei), while the fragile Li, Be and B isotopes are extremely less abundant. However, at a global level, the light H and He are overwhelmingly more abundant than the more strongly bound C, N and O, which in turn are more abundant than the even more stable Fe-peak nuclei.

The (local) correlation between cosmic abundances and nuclear stability suggests that nuclear reactions have shaped the abundances of elements in the universe. The fact that the correlation is only local and does not hold at a global level, tells us that nuclear processes have affected only a small fraction of the baryonic matter in the universe (less than a few per cent). This is good news, since the overabundant H and He nuclei constitute the main fuel of stars (see 3). The history of stars and of the associated nuclear transmutations is far from its end yet.

1.2.3 Overview of Nucleosynthesis

Based on a rapidly growing body of empirical data, both astronomical (abundances in stars and meteorites) and nuclear (binding energies and nuclear reaction rates), as well as on an elementary understanding of stellar structure and evolution, Burbidge et al. (1957)⁵ and Cameron (1957) identified in two landmark papers the main nucleosynthetic processes in nature. These processes have been thoroughly studied throughout the second half of the twentieth century. It is now well established that:

- The light isotopes of H and He, along with 10% of the fragile ${}^7\text{Li}$, have been produced in the hot early universe by thermonuclear reactions between neutrons and protons.⁶

⁵ One of the authors, W.A. Fowler, received the 1984 Nobel Prize in Physics for his contribution to our understanding of the origin of the elements.

⁶ Note that nuclei with mass numbers $A = 5$ and 8 are unstable and do not exist in nature. This fact has important implications for primordial nucleosynthesis, since

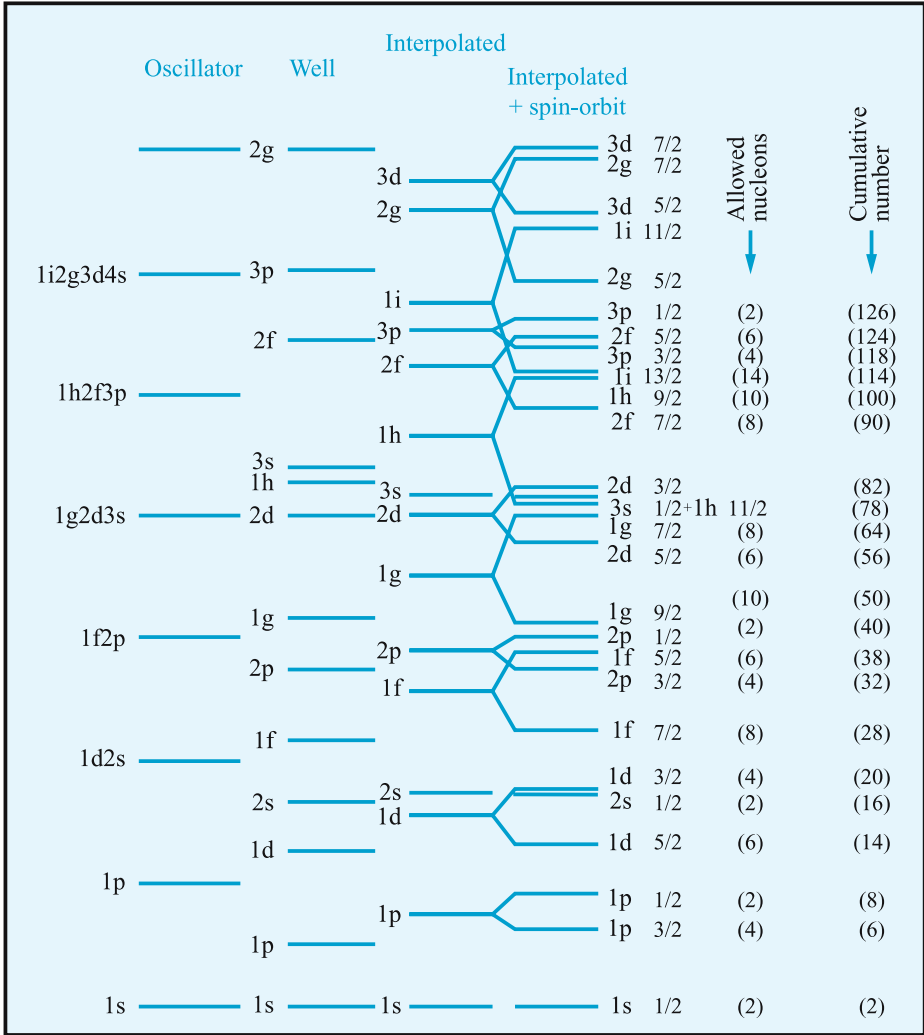


Fig. 1.4. Nuclear level scheme for various nuclear potentials. The spin-orbit interaction (*fourth column*) was of fundamental importance for a correct understanding of nuclear properties. Nuclei with filled shells (with 2, 8, 20, 28, 50, 82 and 126 nucleons, *last column*) are considerably more stable than their neighbors, a property also reflected in the cosmic abundance curve

two body reactions could not bridge the gap between ${}^4\text{He}$ and ${}^{12}\text{C}$. Also, because of the rapid decrease of density and temperature in the hot early universe, the $3\alpha \rightarrow {}^{12}\text{C}$ reaction had no time to operate either; this became possible only much later, inside red giant stars (see Sect. 1.3.3).

- All the elements between C and the Fe peak have been produced by thermonuclear reactions inside stars, either during their quiescent evolutionary stages, or during the violent explosions (supernovae) that mark the deaths of some stars.
- Elements heavier than those of the Fe peak have been produced by neutron captures in stars (see Sect. 1.6), either in low neutron densities and long timescales (s-elements) or in high neutron densities and short timescales (r-elements); a minor fraction of those heavy elements has been produced by photodisintegration of the heavy isotopes in supernova explosions (p-isotopes).
- Finally, the light and fragile isotopes of Li, Be and B are not produced in stellar interiors (they are rather destroyed in high temperatures), but by *spallation reactions*, with high energy cosmic ray particles removing nucleons from the abundant C, N and O nuclei of the interstellar medium.

In the following we shall focus on the stellar production of heavy nuclei (metals), which are much more important than the light nuclei for the origin of terrestrial planets and of life. This will require a rapid tour through the properties of stars and, in particular, of their interiors.

1.3 Stars: from the Main Sequence to Red Giants

1.3.1 Basic Stellar Properties

The theory of stellar structure and evolution is arguably the most successful theory in the whole of astrophysics. It relies heavily on the interpretation of the famous Hertzsprung–Russell diagram (H–R diagram), established by E. Hertzsprung and H.N. Russel in the 1910s. This diagram (Fig. 1.5) concerns two fundamental stellar properties: the *absolute luminosity* L (derived from the apparent luminosity, once the distance is known) and the surface temperature (measured through the color of the star and called *effective temperature* T_E when the stellar surface is assumed to radiate like a black body).

In the solar neighborhood, 90% of the stars lie on the main sequence, a quasi-diagonal band running from high to low values of L and T_E , with the more massive stars being hotter and more luminous.⁷ The remaining 10% are either cold (red) and luminous objects (red giants) or hot and subluminous ones (white dwarfs).

The interpretation of the H–R diagram became possible once the nature of the energy source of stars was elucidated. The works of H. Bethe and C.-F. von Weizsaecker in the 1930s established the series of nuclear reactions that produce the energy radiated by the Sun and the other main sequence stars

⁷ Stellar masses can be accurately determined only in binary systems, by application of Kepler’s laws.

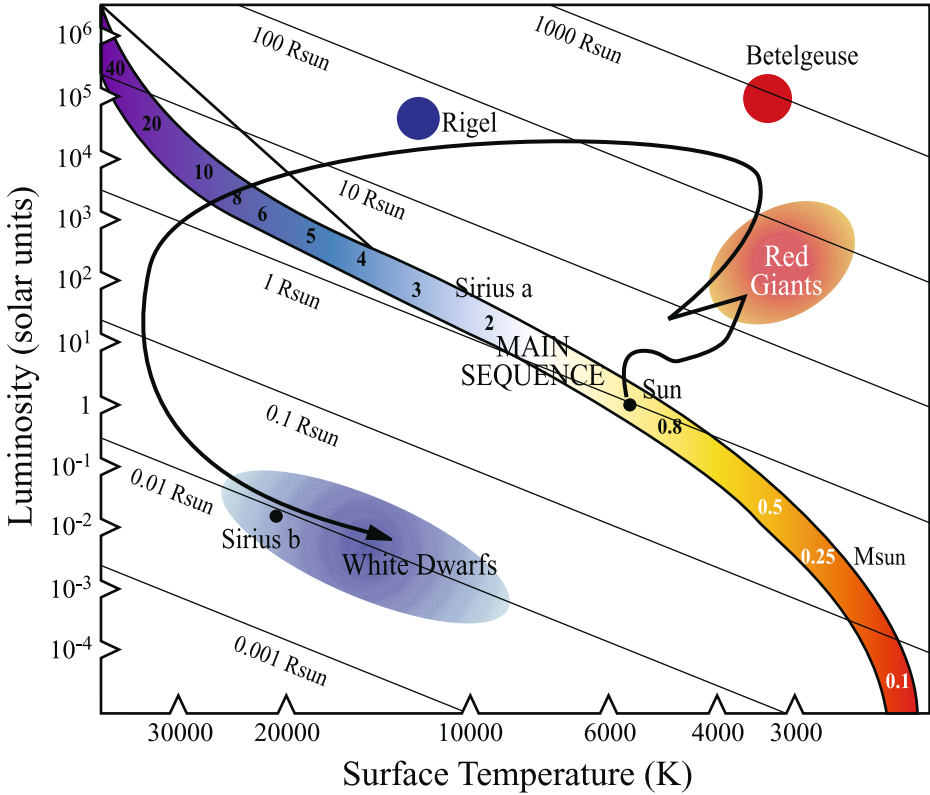


Fig. 1.5. Stellar luminosity vs. effective temperature. On the main sequence (the *quasi*diagonal shaded band, including $\sim 90\%$ of the stars in the solar neighborhood) numbers indicate the corresponding stellar mass (in solar units). *Diagonal lines* indicate stellar radii (in solar units R_{Sun} at a given position of the diagram). The *curve* shows schematically the evolution of the Sun, through the red giant and down to the white dwarf final stage

(see Sect. 1.3.2).⁸ The structure of those stars was rather well understood at that time, after the work of A. Eddington in the 1920s.

To a first (and, actually, quite good) approximation, a star is a gaseous sphere in hydrostatic equilibrium between (1) the attractive force of its own gravity (depending on its mass), and (2) the internal pressure, which depends on the physical state of the stellar gas; for a perfect gas (dominating the interiors of main sequence stars) it is proportional to the product of temperature and density.

⁸ Hans Bethe received the 1967 Nobel Prize in Physics for his work on the energy production in the Sun and stars.

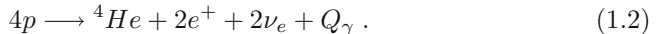
It turns out that gaseous masses of solar composition between 0.08 and 100 solar masses (M_{\odot}) find equilibrium for central temperatures in the range of 2 to 30 million K. This has two important consequences:

1. Because of the temperature gradient between the center and the surface, the star radiates its internal heat.
2. The central temperatures are high enough to induce thermonuclear fusion reactions between the abundant and light hydrogen nuclei, which liberate huge amounts of energy and help keep the stellar interior hot for long durations (see Sect. 1.3.2).

Note that, thanks to its gravity, the star *controls* the rate of the nuclear reactions in its interior and does not explode; indeed, should the nuclear reaction rate increase, producing more energy, the star would heat up, expand⁹ and cool, thus reducing the energy production and coming back to its previous fuel consumption rate. Thus, a star is a *gravitationally bound thermonuclear fusion reactor*; it shines simply because it is massive¹⁰, and for long durations because it uses an efficient energy source, and it has large amounts of fuel available.

1.3.2 H-Burning on the Main Sequence

Depending on internal temperature, H-burning may take place through different modes inside stars. The overall result, however, is always the same: four protons disappear and give rise to a ${}^4\text{He}$ nucleus, while two positrons and two neutrinos are released, as well as γ -ray photons:



The energy released, corresponding to the mass difference Δm between the 4 protons and the ${}^4\text{He}$ nucleus is $E \sim \Delta mc^2 \sim 26\text{MeV}$, i.e., $\sim 6.6\text{MeV/nucleon}$ or 5×10^{18} ergs/g. Most of the energy is deposited locally and heats the stellar gas, while a minor fraction escapes the star, carried away by neutrinos (which interact only weakly with matter). Neutrinos from the Sun have been detected since the 1960s by several experiments. The detected fluxes are in excellent agreement with predictions of solar models, once neutrino oscillations are taken into account (as suggested by the Sudbury neutrino experiment in 2000); this agreement is strong and clear evidence that energy production in the Sun is indeed well understood.

⁹ Because a perfect gas pressure depends on both density *and* temperature, i.e., $P \propto \rho T$; this is not the case for a degenerate gas, and this has explosive consequences for SNIa (see Sect. 1.5.4).

¹⁰ Even without nuclear reactions, the Sun could shine with its current luminosity L_{\odot} for $\sim 30\text{Myr}$, by slowly contracting and releasing gravitational energy $d(GM^2/R)/dt \sim L_{\odot}$. The required contraction rate is $dR/dt \sim 7\text{m/yr}$, too small to be detectable.

In stars of mass $< 1.2 M_{\odot}$ and central temperatures below 20×10^6 K most of the energy is produced by the proton–proton chains (p–p chains) (Fig. 1.6). The first of these reactions involves the conversion of a proton to a neutron, through a weak and very slow interaction, which explains the very long lifetimes of stars powered by this H-burning mode (~ 10 Gyr for the Sun).

In stars of higher masses and temperatures, H-burning occurs through the CNO cycle (Fig. 1.7), where the C, N and O isotopes (produced from previous stellar generations) act as catalysts. The sum of the abundances of CNO nuclei remains constant throughout H-burning, but there is an internal rearrangement: ^{12}C and ^{16}O turn into ^{14}N and, to a smaller extent, into ^{13}C and ^{17}O ; these are the main nuclei produced by the CNO cycle.

An important difference between the p–p chain and the CNO cycle concerns the dependence of the respective energy production rates on temperature. In the case of the p–p chains energy production rate scales as $\epsilon_{\text{pp}} \propto T^4$, while in the case of the CNO cycle it scales as $\epsilon_{\text{CNO}} \propto T^{18}$. This difference is due to the effect of Coulomb barriers between reactants, which are higher in the latter case. The strong temperature dependence of ϵ_{CNO} has an important implication: the energy produced locally by the CNO cycle can only be evacuated by convection, which implies that the stellar interior becomes chemically well-mixed (i.e., nuclide abundances are uniform inside the convective region). On the contrary, in the case of the Sun and low mass stars, energy is evacuated by radiation; the abundances of reacting nuclei and of their products vary smoothly with radius in the interiors of such stars.

On the main sequence the luminosity of a star (i.e., its fuel consumption rate) is proportional to some power of its mass ($L \propto M^K$, with $K \sim 3$ in the upper and $K \sim 4$ in the lower main sequence), while the available energy E is only

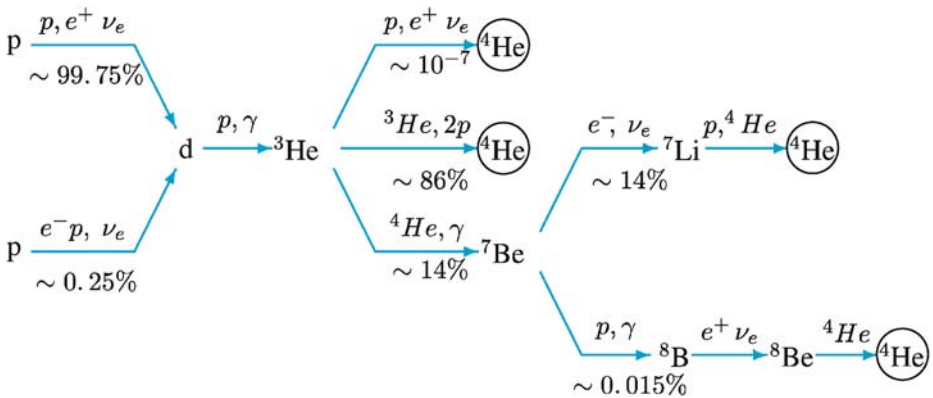


Fig. 1.6. Nuclear reactions of the three proton–proton chains, producing one ^4He nucleus from 4 protons and providing most of the energy of low mass stars on the main sequence. The percentages of occurrence (86% for the first chain, etc.) apply to conditions in the present day solar interior

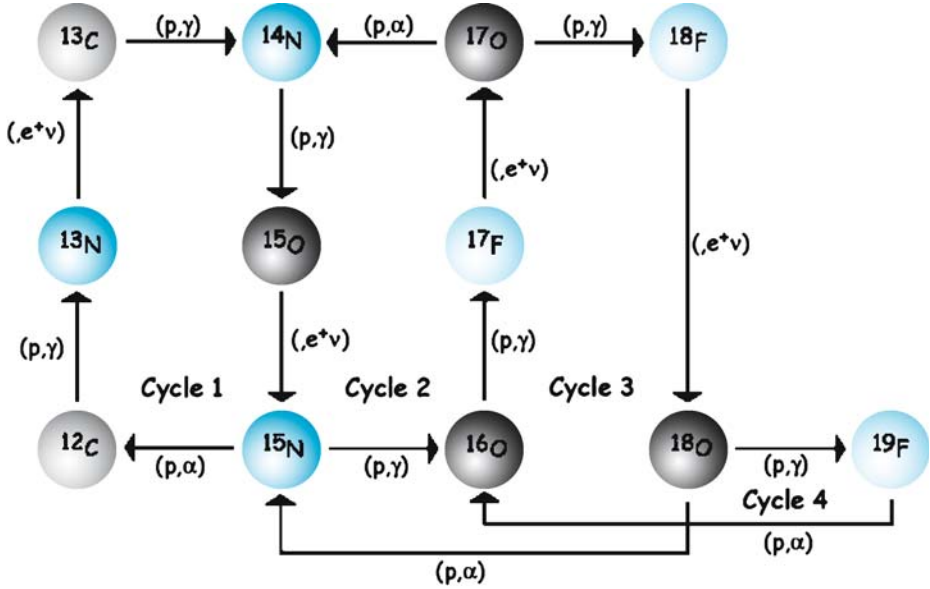


Fig. 1.7. The full CNO tri-cycle, using the C, N and O isotopes as catalysts during H-burning. It provides energy in main sequence stars more massive than $1.2 M_{\odot}$ and produces ^{14}N from the initial ^{12}C and ^{16}O . (Note: Reaction $A + \text{p} \rightarrow B + \text{x}$ can also be written as $A(\text{p}, \text{x})B$; the latter notation is generally adopted in the text, while the former only occasionally)

proportional to the mass (a fraction $f \sim 10-50\%$ of the mass is “burned”, with larger fractions for larger values of M). Thus the lifetime on the main sequence is $\tau \propto fE/L \propto M/M^K \propto M^{-2}$. The most massive stars (several tens of M_{\odot}) shine for only a few megayears, a short lifetime compared to the Sun (10Gyr) or to our current closest neighbor, Proxima Centauri (a $0.12 M_{\odot}$ star, bound to live for more than 1000Gyr).

Note that at H-exhaustion the star has *used* $\sim 6.6\text{MeV/nucleon}$ out of the $\sim 8.8\text{MeV/nucleon}$ available, i.e., before reaching the ultimate nuclear stability in the Fe peak (see Fig. 1.3 bottom). In other terms, it has already spent the largest part of its nuclear fuel. This explains why the H-burning phase is the longest period in a star’s life and why most of the stars are found on the main sequence.

1.3.3 He-Burning in Red Giants

After H-exhaustion, the stellar core contracts and releases gravitational energy, which brings the surrounding H-rich layers to temperatures high enough for H-burning reactions. At the same time the envelope expands and cools and the star turns into a red giant (or supergiant for the most massive of them, i.e.,

above $10 M_{\odot}$). The envelope of such a star is convective and brings to the surface products of the previous central H-burning phase. Enhanced abundances of ${}^4\text{He}$ and ${}^{14}\text{N}$, as well as modified isotopic ratios of, e.g., ${}^{13}\text{C}/{}^{12}\text{C}$ or ${}^{17}\text{O}/{}^{16}\text{O}$, are the marks of this “first dredge-up” phase, which allows us to compare nucleosynthesis theory to observations.

When the H-exhausted core reaches temperatures of $100\text{--}200 \times 10^6$ K (with higher temperatures corresponding to higher stellar masses), He fusion begins. It proceeds in two steps (see Fig. 1.8), with the second one involving a *resonant* reaction.¹¹ The final outcome is the formation of a ${}^{12}\text{C}$ nucleus from three α particles (3α reaction).

The fusion of ${}^4\text{He}$ to ${}^{12}\text{C}$ is much less energetically efficient than the fusion of H to He. It releases an energy of $[3m(\alpha) - m({}^{12}\text{C})]c^2/\sim 7.3\text{MeV}$ or $\sim 0.6\text{MeV/nucleon}$, i.e., about 10 times less energy per unit mass than H-burning. This explains why the number of red giants is so much smaller than the number of main sequence stars and constitutes another important test of the theory of stellar evolution.

During He-burning, ${}^{12}\text{C}$ nuclei capture α particles to form ${}^{16}\text{O}$ nuclei; the ${}^{12}\text{C}(\alpha, \gamma){}^{16}\text{O}$ reaction is not resonant, however, so that at the end of He-burning a considerable amount of ${}^{12}\text{C}$ is left over in the stellar core. ${}^{16}\text{O}$ is usually dominant, but the exact ${}^{16}\text{O}/{}^{12}\text{C}$ ratio depends on the rate of the ${}^{12}\text{C}(\alpha, \gamma){}^{16}\text{O}$ reaction, which is still uncertain.

Note also that in the very beginning of He-burning, at temperatures $\sim 100 \times 10^6$ K, ${}^{14}\text{N}$ turns into ${}^{18}\text{O}$ and later into ${}^{22}\text{Ne}$, through successive α captures. Thus, He-burning constitutes the production mode of ${}^{18}\text{O}$ in the Universe, since some amount of it survives in the He-shell. Note also that, towards the end of He-burning in massive stars, at temperatures $T \sim 250 \times 10^6$ K, neutrons are released through ${}^{22}\text{Ne}(\alpha, n){}^{25}\text{Mg}$. Due to the lack of Coulomb barriers, those neutrons are easily captured by all nuclei in the star (in proportion to the corresponding neutron capture cross-sections). Part of those nuclei are destroyed in subsequent stages of the evolution of massive stars, but those that survive are ejected by the final supernova explosion. This weak s-process leads to the production of the light s-nuclei, with mass number A between 60 and 90, in nature (see also Sect. 1.6.2).

The most massive stars (with initial masses above $30 M_{\odot}$) develop strong stellar winds, due to the high radiation pressure on their envelopes. Losing more than $10^{-5} M_{\odot}$ per year, they finally reveal layers that have been affected by nucleosynthesis. The products of core H-burning (${}^4\text{He}$ and ${}^{14}\text{N}$) and then of core He-burning (${}^{12}\text{C}$, ${}^{16}\text{O}$ and ${}^{22}\text{Ne}$) appear thus with enhanced abundances on the stellar surface. Those stars are called *Wolf-Rayet stars* and their spectroscopy

¹¹ The existence of the 7654 keV level in the nucleus of ${}^{12}\text{C}$, which renders the 3α reaction resonant, was *predicted* by Hoyle (and confirmed by experimenters) in 1953; Hoyle argued that ${}^{12}\text{C}$ in the Universe *should be made* by the 3α reaction in red giant conditions.

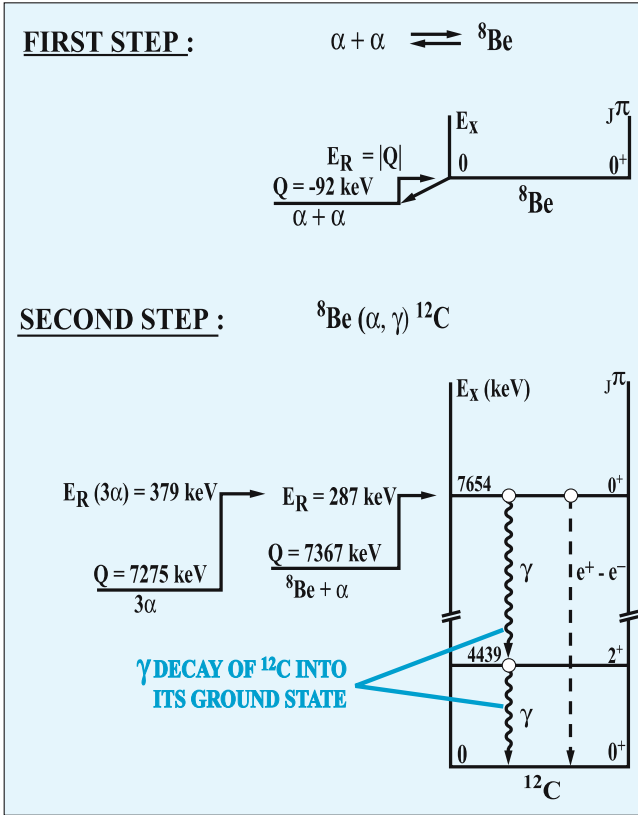


Fig. 1.8. The triple-alpha (3α) reaction occurs in two steps: formation of ${}^8\text{Be}$ from two alpha particles (*top*) and capture of a third alpha by ${}^8\text{Be}$ during the 10^{-16} s of the lifetime of this unstable nucleus (*bottom*); at characteristic red giant temperatures, the second step becomes rapid enough only because of the existence of the level at 7654 keV of the ${}^{12}\text{C}$ nucleus, which makes the reaction resonant (adopted from Rolfs and Rodney 1987)

offers an invaluable test of our nucleosynthesis theories (Maeder and Meynet 2003, and references therein).

After core He-exhaustion the star proceeds in a way similar to the “after H-burning”: the carbon-oxygen core contracts and ${}^4\text{He}$ ignites in the surrounding layers, while H may also burn in even more external layers. In the case of stars with masses $M < 8 M_\odot$, material from the He-layer is mixed first in the H-layer and finally in the stellar envelope, which becomes enriched in He-burning products after this “3d dredge-up” phase. The existence of “carbon stars”, i.e., of low mass red giants with high carbon abundances, offers another important test of the theory of stellar evolution and nucleosynthesis.

In the range of intermediate and low mass stars ($M < 8 M_{\odot}$) the double-shell burning phase is highly unstable; energy is released in “thermal pulses” which progressively expel the stellar envelope into space. The star becomes a planetary nebula for a few tens of thousands of years. After that, the naked C-O core slowly cools down. Its temperature never rises to the point of carbon ignition, since the pressure of its degenerate electron gas can resist its gravity forever. The star becomes less and less luminous until it ends its life in the “graveyard” of white dwarfs (see Fig. 1.5). The fate of stars more massive than $10 M_{\odot}$ is quite different (and much more spectacular).

1.4 Advanced Evolution of Massive Stars

Despite their small number (less than 1% of a stellar generation) massive stars constitute the most important agents of galactic chemical evolution. Indeed, most of the *heavy* elements (metals) in the universe are synthesized in the hot interiors of massive stars, and in particular during the final supernova explosion.

Contrary to their lower mass counterparts, which stop evolving after burning He in a shell surrounding an inert carbon-oxygen core, stars with mass $M > 10 M_{\odot}$ burn successively all the available nuclear fuels in their core, until its composition is dominated by nuclei of the iron peak. However, the duration of all stages subsequent to core He-burning is so short that no direct observational tests of the evolutionary status of the core (which is hidden inside a red supergiant envelope) are possible. Only *post-mortem* observations offer a possibility of (indirectly) validating the results of our models.

1.4.1 Neutrino Losses Accelerate Stellar Evolution

One of the most important features of the advanced evolutionary stages of massive stars is the copious production of neutrinos, due to the high temperatures and densities reached in the stellar core after He-exhaustion ($T > 0.8 \times 10^9 \text{K}$, $\rho > 10^5 \text{g cm}^{-3}$). Neutrino-antineutrino pairs are produced by three processes:

1. *Electron-positron annihilation.* The e^- s and e^+ s are created by the hot thermal plasma (note that only a small fraction of the annihilations leads to $\nu - \bar{\nu}$ pair production).
2. *Photo-neutrino process.* This process is analogous to Compton scattering, with the outgoing photon replaced by a $\nu - \bar{\nu}$ pair.
3. *Plasma process.* A process where a “plasmon” (an excitation of the plasma with an energy $\hbar\omega_p$, where ω_p is the plasma frequency) decays into a $\nu - \bar{\nu}$ pair.

Neutrinos interact only weakly with matter, with a cross-section $\sigma_{\nu} \sim 10^{44} \text{cm}^{-2}$. Their mean free path in the stellar core, at densities $\rho \sim 10^5 \text{g cm}^{-3}$ at carbon ignition (see Sect. 1.4.2) is $l_{\nu} \sim (\rho N_A \sigma_{\nu})^{-1} \sim 10^5 R_{\odot}$ ($N_A = 6.023 \times 10^{23}$

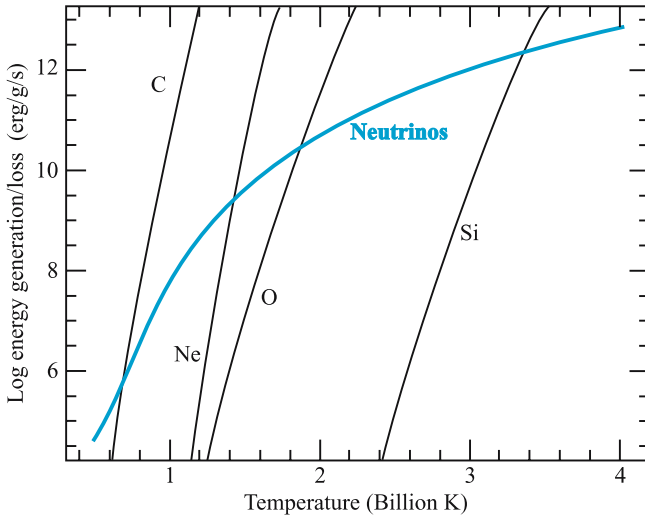


Fig. 1.9. Energy production rate (per unit mass) during the advanced evolutionary stages of massive stars (C, Ne, O and Si-burning) and neutrino loss rates as a function of temperature. Actual burning temperatures for each stage are found at the intersections of the neutrino loss curve with the corresponding energy production curve (from Woosley, Heger and Weaver 2002)

being Avogadro's number). Neutrinos escape then from the stellar core, taking away most of the available thermal energy (produced by nuclear reactions and/or gravitational contraction). In order to compensate for that neutrino *hemorrhage* the core has to increase its nuclear energy production, by contracting and increasing its temperature. As a result, the evolution of the star is greatly accelerated: the time between C-ignition and the final supernova explosion is less than 0.1% of the corresponding main sequence lifetime.

The extreme sensitivity of nuclear energy production and neutrino losses on temperature allows one to obtain a (relatively) good estimate of the burning temperature of each nuclear burning stage: the so-called balanced power approximation (Woosley, Arnett and Clayton, 1973) states that the local power production $\dot{\epsilon}_{\text{nuc}}$ just equals the neutrino loss rate $\dot{\epsilon}_{\nu}$ (a dot over a symbol indicates a time derivative). Taking into account that both rates depend on temperature and density, one can find the burning temperature by assuming some appropriate density.

By applying the balanced power approximation the burning temperatures of C, Ne, O and Si are derived in Woosley (1986) and are presented in graphical form in Fig. 1.9. One sees that C burns at $T_9 \sim 0.85$, Ne burns at $T_9 \sim 1.4$, O at $T_9 \sim 1.8$ and Si at $T_9 \sim 3.4$ (where T_9 is the temperature in 10^9 K). The corresponding densities are $\sim 10^5$ g cm $^{-3}$ for C-burning, $\sim 10^6$ g cm $^{-3}$ for Ne-burning, a few 10^6 g cm $^{-3}$ for O-burning, and a few 10^7 g cm $^{-3}$ for Si-burning.

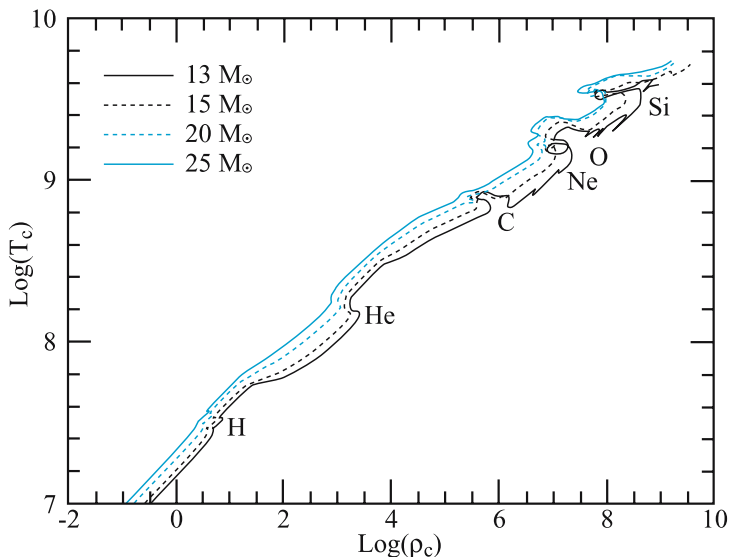


Fig. 1.10. Evolutionary tracks of central temperature (in K) vs. central density (in g cm^{-3}) for stars of 13, 15, 20 and 25 M_{\odot} (from Limongi, Straniero and Chieffi 2000)

The evolution of the central temperature and density for stars in the 13–25 M_{\odot} range (with no mass loss) is shown in Fig. 1.10. Assuming that between two burning phases the stellar core (of mass M and radius R) contracts in quasihydrostatic equilibrium one has for the internal pressure $P \propto M^2/R^4$; combined to the density ($\rho \propto M/R^3$), this leads to $d \ln P = 4/3 d \ln \rho$. Assuming that the core material is in the perfect gas regime ($d \ln P = d \ln \rho + d \ln T$) one finally obtains $T \propto \rho^{1/3}$. The numerical results displayed in Fig. 1.10 follow that relation up to the onset of copious neutrino emission (slightly preceding C-ignition in the core). In fact, after Ne-burning the very center of the star is in the partially degenerate regime, but in the largest part of the burning core conditions are still those of the perfect gas. Note that during periods of nuclear energy production, the relation $T \propto \rho^{1/3}$ is not satisfied anymore (the curves turn slightly to the left).

1.4.2 C, Ne, and O-Burning

Carbon burning in massive stars occurs at $T_9 \sim 0.8$ and $\rho \sim 10^5 \text{ g cm}^{-3}$. The core composition at C-ignition is dominated by the ashes of He-burning, ^{12}C and ^{16}O (more than 90% of the total) in a proportion that decreases with the stellar mass. The exact proportion of $^{12}\text{C}/^{16}\text{O}$ and the exact initial fraction of ^{12}C (which is crucial for the energetics of C-burning) depend on the (still uncertain) value of the $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ rate and on the adopted criterion of convection during He-burning. The reason for the latter dependence is that every α particle

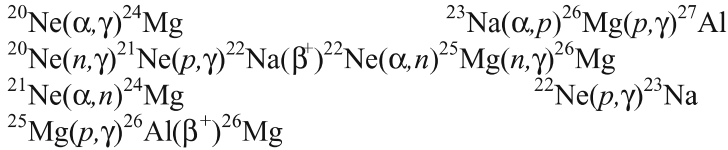
Basic :**Down of 10^{-2} of above :**

Fig. 1.11. Main nuclear reactions taking place during C-burning (the notation $A(x,y)B$ is equivalent to $A+x \rightarrow B+y$). The first four (*top*) reactions produce the quasitotality of the nuclear energy. The others are less important energetically, since their *reaction fluxes* (the product of the abundances of the reactants times the corresponding reaction rate) are smaller than (down to 10^{-2} of) the flux of the first four reactions; however, they contribute to nucleosynthesis during C-burning (from Thielemann and Arnett 1985)

brought in the convective core while He is almost exhausted (and the rate of the 3α reaction too low to produce new ${}^{12}\text{C}$ nuclei) converts a ${}^{12}\text{C}$ nucleus into ${}^{16}\text{O}$.

The fusion of two ${}^{12}\text{C}$ nuclei produces a compound nuclear state of ${}^{24}\text{Mg}$, which decays by emitting a proton or an α particle (see Fig. 1.11). Note that the neutron emission channel, i.e., ${}^{12}\text{C} + {}^{12}\text{C} \rightarrow n + {}^{23}\text{Mg}$, corresponds to an endothermic reaction with a small probability (0.1% at $T_9 = 1$) but it is nevertheless important because the decay of ${}^{23}\text{Mg}$ to ${}^{23}\text{Na}$ changes the neutron excess (η)

$$\eta = \sum_i (N_i - Z_i) Y_i \qquad (1.3)$$

where N_i , Z_i and $Y_i = X_i/A_i$ are the neutron number, the charge, and the number fraction of nucleus i , respectively (see Sect. 1.4.3). Moreover, neutron captures of heavy nuclei produce heavier nuclei than Fe-nuclei, thus modifying the s-process composition resulting from the previous He-burning phase. Also, protons and α particles released by the ${}^{12}\text{C} + {}^{12}\text{C}$ fusion are captured in the ambient nuclei through dozens of (energetically unimportant) reactions. In particular, α captures of ${}^{20}\text{Ne}$ produce ${}^{24}\text{Mg}$. Thus, the main products of C-burning are ${}^{20}\text{Ne}$, ${}^{23}\text{Na}$ and ${}^{24}\text{Mg}$.

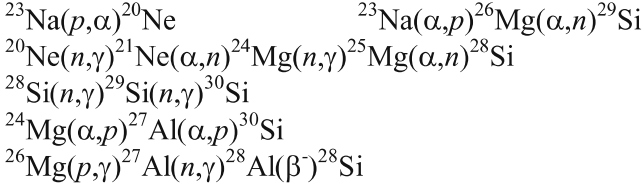
Basic reactions:**Flow > 10² times the above:**

Fig. 1.12. Main nuclear reactions taking place during Ne-burning (from Thielemann and Arnett 1985)

Carbon burning releases an energy $q_{\text{nuc}} \sim 0.4 \text{ MeV/nucleon}$ or $\sim 4 \times 10^{17} \text{ erg/gr}$. The energy production rate (expressed in erg/g/s) is

$$\dot{\epsilon}_{\text{nuc}} \propto Y_{12}^2 \rho \lambda_{12,12} \qquad (1.4)$$

where $\lambda_{12,12}$ is the ${}^{12}\text{C} + {}^{12}\text{C}$ fusion reaction rate and Y_{12} the number fraction of ${}^{12}\text{C}$. For initial number fractions of ${}^{12}\text{C} < 2\%$ (or mass fractions $X = AY_{12} < 0.2$, with $A = 12$) the energy production rate is small enough, so that ${}^{12}\text{C}$ burns radiatively. According to detailed numerical models, radiative burning happens for stars more massive than $\sim 19 M_{\odot}$ (Woosley et al. 2002); in less massive stars, a convective core is formed. Note that the convection criterion also plays a role in the determination of that critical mass.

After C exhaustion, the composition of the stellar core is dominated by ${}^{16}\text{O}$ and ${}^{20}\text{Ne}$ (more than 90% of the total by mass). ${}^{23}\text{Na}$ and ${}^{24}\text{Mg}$ also exist at the few per cent level (in mass fraction). Despite its smaller Coulomb barrier, ${}^{16}\text{O}$ is not the next fuel to burn, since it is exceptionally stable (being a doubly magic nucleus with $Z = N = 8$). The photodisintegration of the ${}^{20}\text{Ne}$ nucleus ${}^{20}\text{Ne}(\gamma, \alpha){}^{16}\text{O}$ becomes energetically feasible at $T_9 \sim 1.5$, i.e., before the fusion temperature of ${}^{16}\text{O}$ nuclei ($T_9 \sim 2$) is reached. The released α particles are captured on both ${}^{16}\text{O}$ (to restore ${}^{20}\text{Ne}$) and ${}^{20}\text{Ne}$ (to form ${}^{24}\text{Mg}$). The net result of the operation is that this ${}^{20}\text{Ne}$ “melting” can be described by:



The photodisintegration of ${}^{20}\text{Ne}$ is endoergic, but the exoergic α captures of ${}^{16}\text{O}$ and ${}^{20}\text{Ne}$ more than compensate for the energy lost. ${}^{20}\text{Ne}$ -burning produces $\sim 0.1 \text{ MeV/nucleon}$ or $1.1 \times 10^{17} \text{ erg/g}$, i.e., about 1/4 the specific energy released by C-burning. Several other energetically unimportant reactions, induced by p, n and α particles occur during Ne-burning; in particular, ${}^{23}\text{Na}$ (a product

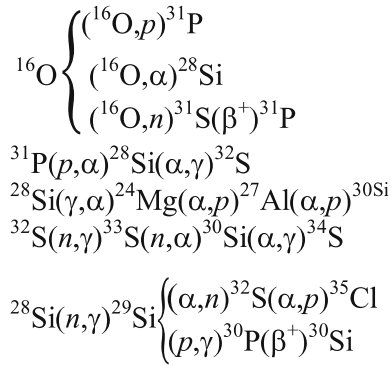
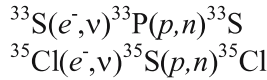
Basic reactions :**Electron captures :**

Fig. 1.13. Main nuclear reactions taking place during O-burning (from Thielemann and Arnett 1985)

of C-burning) disappears through ${}^{23}\text{Na}(p, \alpha) {}^{20}\text{Ne}$ and ${}^{23}\text{Na}(\alpha, p) {}^{26}\text{Mg}$ (see also Fig. 1.12).

After Ne-burning, the stellar core consists mainly of ${}^{16}\text{O}$, ${}^{24}\text{Mg}$ and ${}^{28}\text{Si}$ (the latter being produced mainly through ${}^{24}\text{Mg}(n, \gamma) {}^{25}\text{Mg}(\alpha, n) {}^{28}\text{Si}$). Also, ${}^{29}\text{Si}$, ${}^{30}\text{Si}$ and ${}^{32}\text{S}$ are present at the 10^{-2} level (in mass fraction).

The fusion of two ${}^{16}\text{O}$ nuclei produces a compound nucleus of ${}^{32}\text{S}$, which decays through the p , α and n channels (see Fig. 1.13); the corresponding branching ratios and energy released are 58% (7.68 MeV), 36% (9.58 MeV) and 6% (1.45 MeV), respectively. Note that, at high temperatures, the endoergic decay through the deuteron channel is also effective.

The energy released by oxygen burning is 5×10^{17} erg/g or 0.5 MeV/nucleon. As in the previous stages, dozens of n , p and α induced reactions occur. However, the increased temperature and density introduce two novel features:

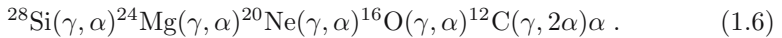
1. Electron captures occur, mainly on ${}^{31}\text{S}$, ${}^{30}\text{P}$, ${}^{33}\text{S}$, ${}^{33}\text{Cl}$ and ${}^{37}\text{Ar}$. These weak interactions modify substantially the neutron excess η , up to $\eta \sim 0.01$ (especially in the lower mass and denser stars); the final neutron-rich composition is clearly non-solar and should be rarely ejected by the star.
2. Photodisintegration reactions become also important and destroy most of the nuclei heavier than Fe that have been built through n captures in the previous burning phases (mostly by s-process during He-burning). However,

this happens only in the innermost and hottest regions of the stellar core; outside them, products of previous burning stages survive.

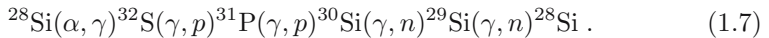
1.4.3 Si-Melting and Nuclear Statistical Equilibrium (NSE)

At O-exhaustion, the composition of the stellar core is dominated by ^{28}Si (30–40% by mass fraction) and, either ^{32}S and ^{38}Ar (in the more massive cores, where low densities do not favor electron captures) or ^{30}Si and ^{34}S (in the lower mass stars, below $\sim 15 M_{\odot}$).

^{28}Si *burns* at a temperature of $T_9 \sim 3.2$ (see Sect. 1.4.1) and in a way that is reminiscent of the ^{20}Ne -burning. The fusion of two ^{28}Si nuclei requires such high temperatures that photodisintegration of all nuclei would result. Instead, part of the ^{28}Si nuclei photodisintegrates, mainly through a sequence of reactions involving α particles:



Other photodisintegration reactions releasing p and n also occur, especially in material with neutron excess η substantially different from zero (due to previous electron captures, i.e., in high density stellar cores). The released α particles and nucleons are further captured by ^{28}Si and heavier nuclei and an equilibrium is established between direct and inverse reactions, e.g.,



The mean atomic weight of the mixture progressively increases, since free nucleons and α particles are held tightly to heavier and more strongly bound nuclei (see Fig. 1.14); that is why the overall process is better described as Si-melting. In fact, local equilibrium between a few nuclear species is established already by the end of O-burning, but, as the temperature increases, the various quasiequilibrium clusters merge. In early Si-melting, there are already two major quasiequilibrium clusters established: one with nuclei with $A = 24$ to 46 and another with Fe-peak nuclei. Towards the end of Si-melting the two clusters merge and the quasiequilibrium group includes all nuclei above ^{16}O . Note that such reaction sequences also bring into equilibrium the abundances of α particles and free nucleons *indirectly* (i.e., not through photodisintegration of α particles in free nucleons). Assuming that two ^{28}Si nuclei turn into one nucleus of the Fe peak, one finds that the energy released by Si-melting is $\sim 0.2\text{MeV/nucleon}$ or $\sim 2 \times 10^{17} \text{ erg/g}$.

Towards the end of Si-melting all electromagnetic and strong nuclear interactions are in equilibrium with their inverses. Since neutrinos still escape freely from the stellar core, neutrino producing weak interactions never come into equilibrium with their inverses and total equilibrium is not established. The last reactions to reach equilibrium are those linking ^{24}Mg to ^{20}Ne , ^{16}O to ^{12}C and,

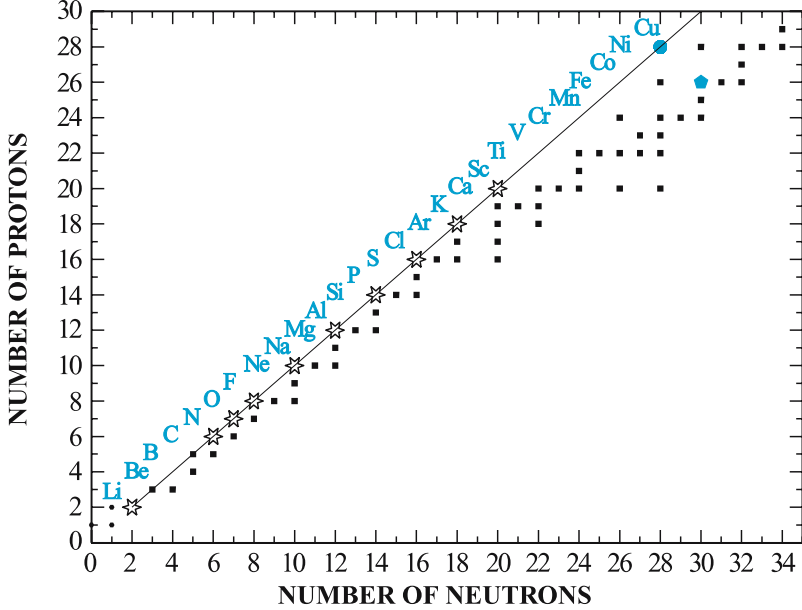


Fig. 1.14. Position of stable nuclei on the proton–neutron diagram. Stellar nucleosynthesis proceeds at $Z = N$ up to ^{40}Ca , favoring the strongly bound α nuclei (*asterisks*, see also Fig. 1.3). Weak interactions (electron captures) during late O-burning and Si-melting modify the neutron excess and progressively shift the composition towards ^{56}Fe ($Z = 26$, $N = 30$, *filled pentagon*). In explosive burning conditions, weak interactions have no time to operate and the nuclear flow proceeds at constant neutron excess and $Z = N$ (*solid line*) all the way up to ^{56}Ni ($Z = N = 28$, *filled dot*)

finally, the reaction $3\alpha \longleftrightarrow ^{12}\text{C}$. When this happens, and assuming that all nuclides obey the statistics of an ideal Maxwell–Boltzmann gas, their abundances can be expressed by application of the Saha equation:

$$Y_i(A_i, Z_i) = (\rho N_A)^{A_i-1} \frac{G_i}{2^{A_i}} A_i^{3/2} \left(\frac{2\pi\hbar^2}{m_H kT} \right)^{3(A_i-1)/2} \times e^{B(A_i, Z_i)/kT} Y_p^Z Y_n^{(A-Z)} \quad (1.8)$$

where $G_i = \sum_j (2J_j^i + 1) \exp(-E_j^i/kT)$ is the partition function of nucleus i (a summation over all nuclear levels of spin J and energy E_j) and $B(A_i, Z_i)$, its binding energy. Two more equations are required to eliminate the neutron and proton abundances Y_n and Y_p and these are taken to be the mass conservation equation

$$\sum Y_i A_i = 1 \quad (1.9)$$

and the charge conservation

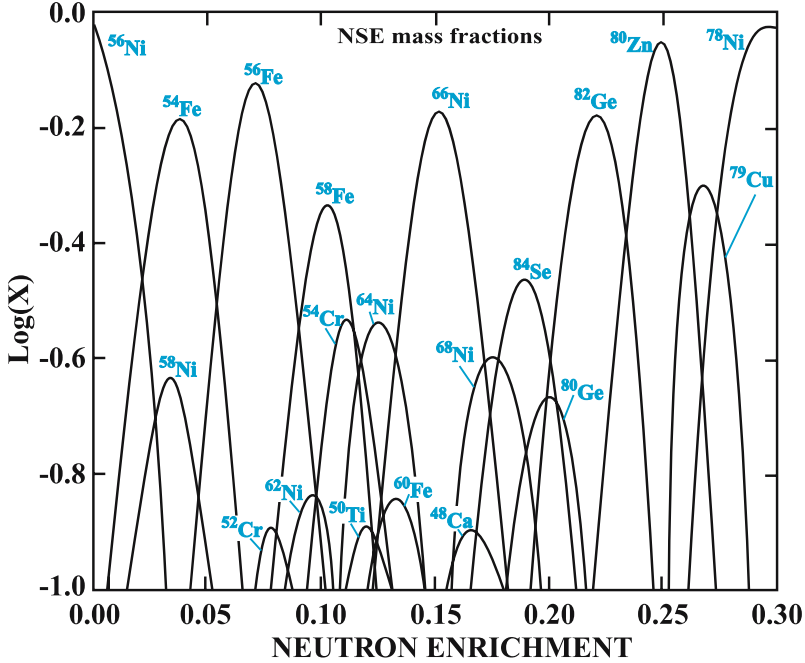


Fig. 1.15. Composition at nuclear statistical equilibrium as a function of the neutron enrichment η ; composition corresponding to $\eta > 0.1$ is not encountered in the cosmic abundance curve (from Hartmann, Woosley and El Eid 1985)

$$Y_e = \frac{n_e}{\rho N_A} = \sum_i Z_i Y_i = \frac{1 - \eta}{2} \quad (1.10)$$

where Y_e is the electron mole number and n_e the electron density. Thus, at nuclear statistical equilibrium (NSE) the composition is described as a function of three parameters: temperature T , density ρ and electron fraction Y_e or neutron excess η . In fact Y_e and η are slowly modified under the action of weak interactions (electron captures) and that variation must also to be taken into account, e.g., through

$$\frac{dY_e}{dt} = \sum_i -\lambda_i Y_i \quad (1.11)$$

where λ_i is the weak interaction rate of nucleus i .

In the conditions of NSE, the Saha equation implies that for not too high of temperatures ($T_9 < 10$) the most abundant nuclei are those with the largest binding energy for a given value of η . Thus, for $\eta \sim 0$, the most tightly bound nucleus is ^{56}Ni (Fig. 1.15). Indeed, in explosive nucleosynthesis ^{56}Ni is the dominant product of NSE; its radioactive decay has been (indirectly) observed in the

case of the supernova SN1987A (see Sect. 1.5.2) brilliantly confirming the theory. However, in the quiescent burning conditions of late stellar evolution, weak interactions have time enough to modify η to values $\sim 0.06\text{--}0.08$, corresponding to $Y_e \sim 0.46\text{--}0.47$. In those conditions, the most tightly bound nucleus is the stable ^{56}Fe , while ^{52}Cr is also produced in substantial amounts (see Fig. 1.15).

When the composition of the stellar core becomes dominated by the strongly bound nuclei of the Fe peak, the star has arrived at the end of its quiescent life. Further contraction of the core and increase of its temperature cannot release nuclear energy to sustain internal pressure: reactions involving stable nuclei at the peak of the binding energy curve are endothermic and constitute sinks of thermal energy. Moreover, electron captures proceed at even higher rates and continue removing electrons from the stellar plasma, weakening the main source of resistance against gravity, namely electron pressure. The iron core collapses. The rate and the details of the collapse depend sensitively on the electron fraction Y_e , since the Chandrasekhar mass (i.e., the maximum mass supported against gravity by a gas of degenerate electrons) is given by

$$M_{\text{Ch}} = 3(8\pi^2)^{1/4} \left(\frac{\hbar c}{2G} \right)^{3/2} (N_{\text{A}} Y_e)^2 \sim 1.45(2Y_e)^2 M_{\odot}. \quad (1.12)$$

Obviously, the rates of electron captures on heavy nuclei play an important role in determining the final value of Y_e before collapse and the fate of the stellar core. Modern calculations produce values around $Y_e \sim 0.445$.

1.4.4 Overview of the Advanced Evolutionary Phases

A schematic view of the interior evolution of a massive star is presented in Fig. 1.16. The life of a massive star is essentially a series of relatively long quiescent central burning stages interrupted by much shorter periods of core contraction and heating, which lead to the ignition of the next fuel. Shell burning of a given nuclear fuel also takes place at the border of the former convective core, where that fuel has been exhausted; the burning shell progressively migrates outwards. However, shell burning occurs at temperatures higher than the corresponding central burning stage and the fuel never burns to exhaustion (incomplete burning).

In practice, the situation is much more complicated than the schematic view of Fig. 1.16. Core burning is not always convective, since it depends on the amount of available fuel, which determines the nuclear energy production (e.g., central C-burning and Ne-burning may occur radiatively). Also, the various burning shells do not burn steadily, but intermittently. Only full scale numerical calculations, including detailed nuclear reaction networks and appropriate physical ingredients can (hopefully) describe “realistically” the advanced stages of massive star evolution (e.g., Arnett 1996, Woosley and Weaver 1995, Thielemann et al. 1996, Chieffi et al. 1998). Note that several key ingredients, like the various mixing processes and the role of rotation, remain still uncertain and may

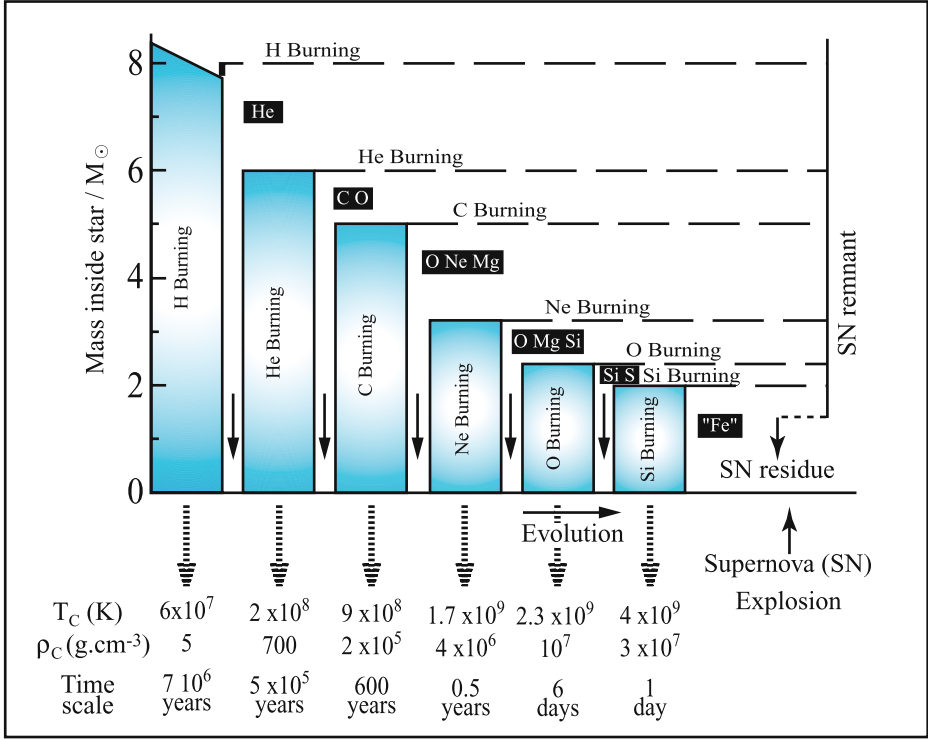


Fig. 1.16. Schematic view of the interior evolution of a $25 M_{\odot}$ star. *Blue regions* indicate convective core burning and *dashed lines* shell burning, while the corresponding main products are inside *boxes*. *Vertical arrows* indicate core contraction between two burning stages. Corresponding central temperatures, densities and burning timescales (approximate values) are given at the *bottom* (from Arnould and Takahashi 1999)

hold important clues as to the fate of stars of all masses (Maeder and Meynet 2002, Heger et al. 2003). Note also that the numerical treatment of oxygen and silicon burning, where convective and nuclear timescales are comparable, is problematic at present (see Bazan and Arnett 1994).

Despite those uncertainties, it is believed that at the end of its quiescent life a massive star develops an “onion-skin” structure, with heavier nuclei dominating the composition as one moves from the surface to the center. The star retains in its various layers the “memory” of its previous burning stages (i.e., in a way analogous – in fact, opposite – to the various geological strata, which keep the “archives” of the earlier periods in the Earth’s history). The reason for that layered structure of the star is the increasing sensitivity of nuclear reaction rates to temperature, as the Coulomb barriers of the successive nuclear fuels increase: the nuclear energy is produced in a region more and more confined to the center of the star, so that the convective cores are smaller at each new burning stage

(shell burning does not modify the picture, since it slightly increases the sizes of all the former convective cores).

With no more nuclear fuel available, the Fe core collapses and the star dies in a spectacular supernova explosion.

1.5 Explosive Nucleosynthesis in Supernovae

1.5.1 Main Properties and Classification of Supernovae

Several supernovae have been observed in historical times in our galaxy by the naked eye (e.g., in 1054 in the constellation of the Crab by Chinese astronomers, in 1572 by Tycho Brahe, in 1604 by Kepler), but none in the past four centuries, i.e., after the invention of the telescope. Dozens of supernovae are routinely observed each year outside the Milky Way, most of them in distances of tens of millions of light-years.

The peak absolute luminosity of a supernova (SN) is 10^{42} – 10^{43} erg/s, i.e., comparable to that of an entire galaxy, while the time-integrated energy that is radiated away is $\sim 10^{49}$ erg, comparable to that radiated by the Sun in 10^8 yr. A hundred times more energy is released in the form of kinetic energy of the ejecta, which are expelled at several 10^3 km/s (and up to a few 10^4 km/s in some cases).

Supernovae are classified on the basis of their spectra (Fig. 1.17) as belonging to type I (absence of characteristic H lines) or type II (presence of H). Despite its interest (few stars show no H in their spectra), this classification concerns only the *epidermic* properties of SN and tells us very little about the mechanism of the explosion.

SNIa are a subtype of SN distinguished by the presence of Si in their early spectra and quasi-identical peak luminosities, which decline in a characteristically exponential way: first with a period of ~ 6 days and after a few weeks more slowly, with a period of 77 days. This exponential decline of SNIa lightcurves is attributed to the radioactive decay chain $^{56}\text{Ni} \rightarrow ^{56}\text{Co} \rightarrow ^{56}\text{Fe}$; the half-lives of ^{56}Ni and ^{56}Co are 6.1 days and 77 days, respectively.

SNIa are encountered in all types of galaxies, including ellipticals, which have ceased star formation billions of years ago and contain only old populations of low mass stars. All other types of supernovae are encountered in galaxies with currently active star formation, i.e., spirals and irregulars. Combined with the lack of H, that property of SNIa suggests that they originate from a population of old objects that have lost their H envelope. White dwarfs are the obvious candidates. Indeed, the thermonuclear burning of $1 M_{\odot}$ of C-O to ^{56}Ni releases about 10^{51} ergs, sufficient to account for the observed energetics of SNIa.

A different mechanism is at the origin of all other SN types: the gravitational collapse of the Fe core of a massive star, which turns into an explosion in a way that is poorly understood up to now. Assuming that the collapse proceeds to the formation of a compact object (a neutron star with radius $R \sim 10$ km) the

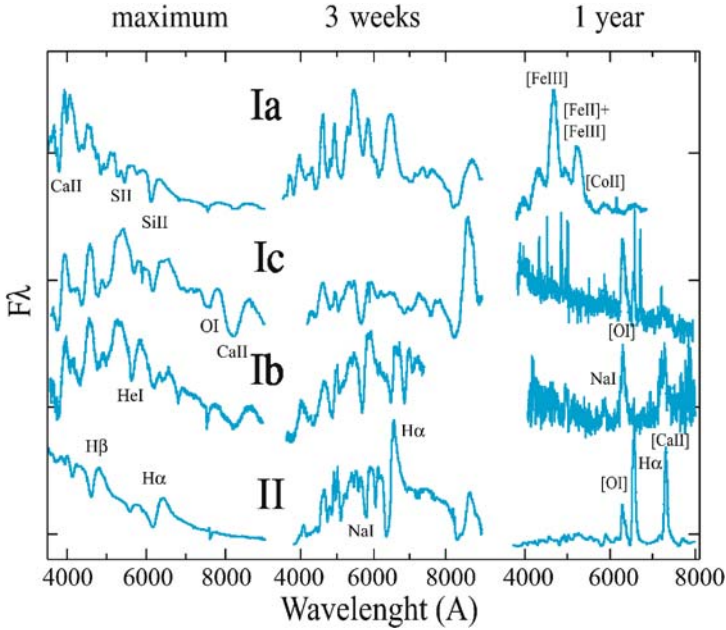


Fig. 1.17. Spectra of SN of various types (from *top to bottom*), taken at three different epochs (from *left to right*)

available gravitational energy of a $1 M_{\odot}$ Fe core ($E \sim GM^2/R \sim 10^{53}$ erg) is more than sufficient to explain the SN energetics. The considerable differences in the properties of those SN (peak luminosities, lightcurves, spectra) are due to the details of the collapse/explosion and, in particular, to the amount of the H-rich envelope left to the star at the moment of its collapse.

From statistics of SN in external galaxies it is estimated that their frequency is $\sim 0.25\text{--}0.35$ SNIa and $1.5\text{--}2$ core collapse SN per century in a galaxy similar in size and type to the Milky Way. The fact that no SN was observed in the past 400yr in our galaxy is attributed to the fact that massive stars are expected to be born and to die mostly in the inner galactic disk (much richer in gas than the outer one), which is hidden from our optical observations by a large column density of gas and dust.

1.5.2 Explosive Nucleosynthesis in Core Collapse Supernovae

The idea that supernova explosions are related to the death of massive stars goes back to the Swiss astronomer F. Zwicky in the 1930s. The first numerical models of such explosions were constructed in the early 1960s (Colgate et al. 1961). However, the mechanism of core collapse supernova explosions remains still highly uncertain (see, e.g., Janka et al. 2002). A schematic view of our current understanding is presented in Fig. 1.18.

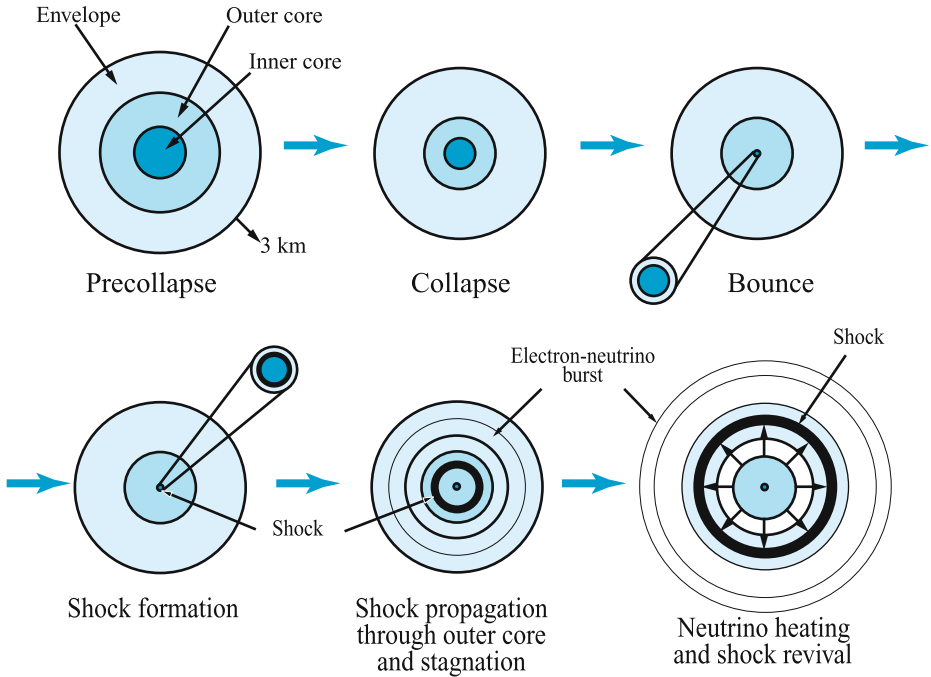


Fig. 1.18. Illustration of the various successive stages of the core collapse of a massive star; see Sect. 1.5.2 for details (from Guidry 1998)

Iron core collapse proceeds in a timescale of milliseconds. Due to increasingly high temperatures, photodisintegrations tear down Fe nuclei to nucleons and alpha particles, while higher densities favor electron captures and conversion of protons to neutrons. When the density of nuclear matter is reached ($\rho \sim 10^{14} \text{ g cm}^{-3}$) the repulsive component of the strong nuclear force brings the collapse of the inner core abruptly to a halt. At the boundary of the inner core a shock wave is formed, with an energy approximately equal to the kinetic energy of the core at the moment of the bounce ($\sim 5 \times 10^{51}$ erg, according to results of numerical simulations). As the shock wave propagates outwards, into the still infalling outer core, it photodisintegrates its Fe nuclei (tearing them down to nucleons) and consequently loses $\sim 8.8 \text{ MeV/nucleon}$ or 1.5×10^{51} erg per $0.1 M_{\odot}$.

If the mass of the infalling outer core is $M_{\text{OC}} = M_{\text{Fe}} - M_{\text{IC}} < 0.4 M_{\odot}$, the shock can reach the base of the envelope with enough energy to launch a successful explosion. Only in the smallest massive stars (below $13 M_{\odot}$, with Fe core mass $M_{\text{Fe}} \sim 1.1 M_{\odot}$) the mechanism of prompt explosion has some chance to succeed. In more massive stars, the prompt explosion fails. One is left with a dense and hot protoneutron star, accreting matter at a rate $1\text{--}10 M_{\odot}/\text{s}$. If

no other energy source is involved, the star collapses to a black hole (at least in most numerical simulations).

It is currently believed that the supplementary energy source is provided by neutrinos, which are “trapped” in the core during the hydrodynamical collapse for a timescale of a few milliseconds¹² and start diffusing outwards after the shock is launched. The protoneutron star radiates its thermal content in the form of neutrinos and in a timescale of a few seconds. These neutrinos carry the largest fraction of the gravitational binding energy of the neutron star ($\sim 3 \times 10^{53}$ erg or $\sim 10\%$ of its rest mass); if they manage to communicate a few percent of that energy in the stalled shock wave, they may induce a delayed explosion. The detection of about 20 neutrinos in the Kamiokande and IBM detectors on February 23, 1987 from supernova SN1987A in the Large Magellanic Cloud demonstrated that neutrinos are indeed copiously produced during the deaths of massive stars.

If the shock wave manages to reach the base of the stellar envelope with enough energy (i.e., a few 10^{51} erg, to account for the gravitational binding energy of the envelope plus the typical kinetic energy of observed SNII ejecta), the explosion is successful. As the shock wave propagates through the stellar envelope, it heats the various layers to temperatures higher than in the corresponding central burning stages. The peak temperature $T_P(r)$ at radius r can be obtained to a good accuracy by assuming that the kinetic energy of the shock KE is equal to the thermal energy in the radiation field behind the shock $E_{\text{TH}} = \frac{4}{3}\pi r^3 a T_P(r)^4$ (where $a = 7.56 \times 10^{-15}$ erg cm⁻³ K⁻⁴ is the blackbody constant). This leads to

$$T_P(r) \sim 1.33 \times 10^{10} \left(\frac{KE}{10^{51} \text{ erg}} \right)^{1/4} \left(\frac{r}{10^8 \text{ cm}} \right)^{-3/4} \text{ K}. \quad (1.13)$$

In Fig. 1.19 it can be seen that this formula gives a very good approximation to the *realistic* peak temperature profile (calculated numerically, right panel of Fig. 1.19). Stellar material in the layers hit by the shock wave is heated to those temperatures for a time on the order of the local hydrodynamical timescale¹³

$$\tau_{\text{HD}} \sim 0.446 \rho_6^{-1/2} \text{ s} \quad (1.14)$$

where ρ_6 is the mean density interior to radius r in 10^6 g cm⁻³.

If the nuclear burning timescale at radius r ($\tau_{\text{nuc}} = q_{\text{nuc}}/\dot{\epsilon}_{\text{nuc}}$, see Sect. 1.4.2) is smaller than the hydrodynamical timescale $\tau_{\text{HD}}(r)$, then explosive nucleosynthesis will take place, modifying the composition left from the previous quiescent

¹² The mean free path of neutrinos becomes temporarily smaller than the dimensions of the collapsing core; however, after 0.1s neutrinos diffuse and escape from the “neutrinosphere”.

¹³ τ_{HD} is essentially the free-fall timescale, i.e., it is assumed that the explosion is symmetric in time to the collapse.

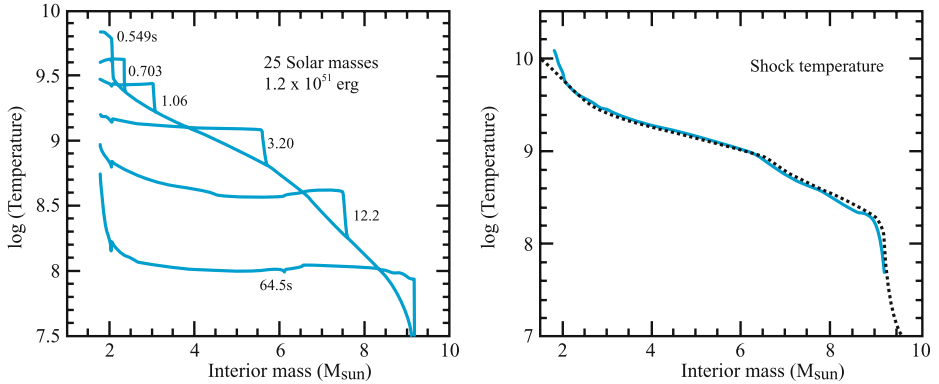


Fig. 1.19. Initial temperature profile (*continuous curve*) and snapshots of the temperature profile at different times after the passage of the shock wave in the mantle of a $25 M_{\odot}$ star (*left*). Peak temperature profile after the passage of the shock wave in the mantle of a $25 M_{\odot}$ star, obtained numerically (*solid curve*) and analytically (*dotted curve*), i.e., by assuming adiabatic expansion of the mantle (*right*) (from Woosley and Weaver 1995)

burning stage. Taking into account the relatively steep density profiles of pre-supernova stars, it turns out that peak temperatures are sufficient for vigorous explosive Si and O-burning; the Ne and C layers are less affected (see Fig. 1.20) while the He and H layers are not affected at all by the explosion (envelope densities are too low for explosive burning to take place in those layers).

The major reactions in explosive burning for a given fuel are the same as those of the corresponding quiescent burning (i.e., Figs. 1.11, 1.12 and 1.13). However, the beta decay lifetimes of unstable nuclei are often larger than the timescales of explosive nucleosynthesis τ_{nuc} . This implies that the cross-sections of nuclear reactions of those unstable nuclei are also required, a situation that is not usually encountered in quiescent burning.

The products of explosive nucleosynthesis in massive stars have been studied in detail in Woosley and Weaver (1995), Thielemann et al. (1996), and Limongi et al. (2000); a recent update can be found in Woosley et al. (2002). In general, one may distinguish three classes of elements/isotopes coming out of the explosion (see Prantzos 2000):

1. The first set (N, C, O, Ne, Mg) results mainly from hydrostatic He and C/Ne-burning (layers not affected by explosive nucleosynthesis). The *yields* (ejected mass) of those elements increase with stellar mass, as do the He and C-exhausted cores. A comparison of those yields to observations (see below) allows to test essentially the physics of the pre-supernova models.
2. The second group (Al, Si, S, Ar, Ca) includes elements produced in both hydrostatic and explosive burning. Their yields vary less with progenitor

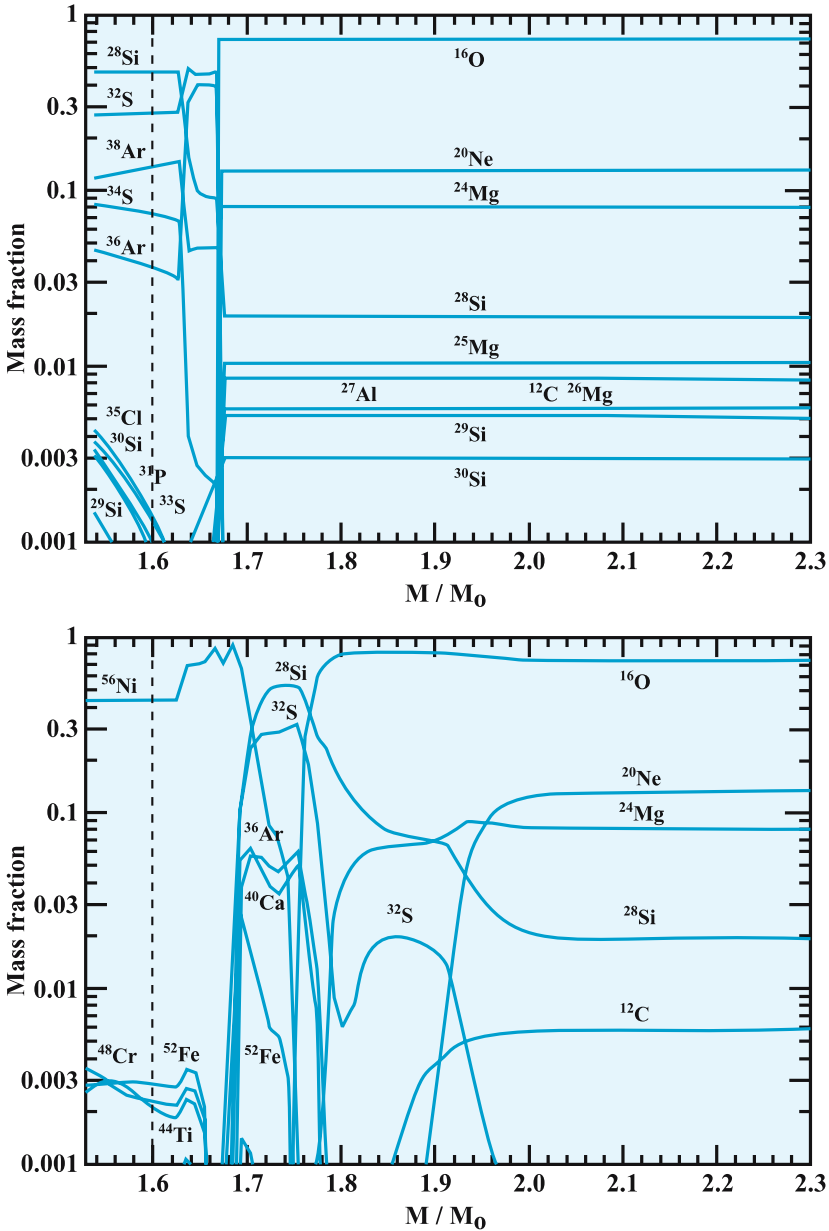


Fig. 1.20. Chemical composition profile of the innermost $2.3 M_{\odot}$ of a $20 M_{\odot}$ star prior to core collapse, with the major products of *hydrostatic* nucleosynthesis; uniform composition (flat abundance profiles) characterize former convective zones (*top*). Chemical composition profile of the inner $2.3 M_{\odot}$ of a $20 M_{\odot}$ star after the passage of the shock wave with the major products of *explosive* nucleosynthesis dominating the innermost $2 M_{\odot}$ (*bottom*) (from Thielemann, Nomoto and Hashimoto 1996)

mass than the ones of the previous group. Comparison to observations tests both pre-supernova models and explosion energy.

3. The third set involves elements of the Fe peak, produced essentially by explosive Si-burning. Their yields are highly uncertain at present, since they depend on the mechanism of the explosion, i.e., the shock wave energy, the various mixing processes and the “mass-cut” (the dividing line between the mass of the star finally ejected and the one that falls back to the compact object).

How can the validity of the theoretical stellar yields be checked? Ideally, individual yields should be compared to abundances measured in supernova remnants of stars with known initial mass and metallicity! However, such opportunities are extremely rare. In the case of SN1987A, theoretical predictions for a $20 M_{\odot}$ progenitor are in rather good agreement with observations of C, O, Si, Cl and Ar (Thielemann et al. 1996). SN1987A allowed also to *calibrate* the Fe yield ($\sim 0.07 M_{\odot}$) from the optical lightcurve (powered at late times from the decay of ^{56}Co , the progeny of ^{56}Ni), extrapolated to the moment of the explosion (e.g., Arnett et al. 1989). More recently, observations of extragalactic supernovae allowed to establish a rather good correlation between the peak luminosity (i.e., the ^{56}Ni mass) and the kinetic energy of the explosion (see Hamuy 2003 for a review).

1.5.3 Explosive Nucleosynthesis in Thermonuclear SN

It is widely accepted now that SNIa are thermonuclear explosions of white dwarfs (WD). Since a single WD is eternally stable, a companion star is required. The currently leading model invokes a carbon-oxygen WD close to the Chandrasekhar mass-limit that accretes mass from a companion star, evolved or not (Fig. 1.21). Another, less popular, scenario involves merging of two WD, after orbital decay (due to energy loss by gravitational wave emission) in a binary system. Although no self-consistent model exists up to now, the former scenario seems to be in much better agreement with observations.

The accretion scenario requires a relatively well-adjusted accretion rate, $\sim 10^{-7} M_{\odot}/\text{yr}$ for several megayears. As the mass of the WD increases, compressional heating rises the internal temperature to several 10^8 K. The high density of the WD ($\sim 10^9 \text{ g/cm}^3$) leads to a very efficient *screening* of the nuclear reactions.¹⁴ $^{12}\text{C} + ^{12}\text{C}$ fusion occurs then in considerably lower temperatures and at much higher rates than in massive stars. Ignition occurs when the $^{12}\text{C} + ^{12}\text{C}$ reaction releases energy faster than neutrino losses (mostly by plasma process, see Sect. 1.4.1) can carry it away.

Contrary to the case of normal stars (where perfect gas dominates the internal pressure), WD are composed of degenerate gas. Its pressure depends only

¹⁴ Because of the high electron density, carbon nuclei feel a much smaller Coulomb repulsion; as a result, their fusion rate is considerably enhanced.

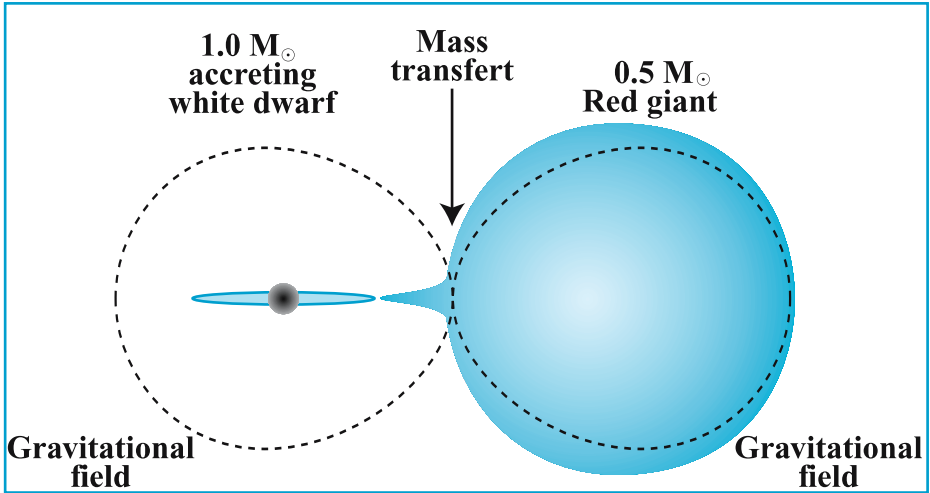


Fig. 1.21. Mass transfer in a binary system, progenitor of a SNIa (personal communication with Timmes, 2005)

on density¹⁵ but not on temperature. Thus, heating of the medium (through the energy released by $^{12}\text{C} + ^{12}\text{C}$) is not accompanied by pressure increase, which would lead to gas expansion and cooling; instead, temperature increases steadily, and so does the $^{12}\text{C} + ^{12}\text{C}$ reaction rate. In those conditions, a carbon deflagration burns explosively the material of the star, all the way towards NSE. The timescale of the explosion is too short to allow for important electron captures, and thus the final outcome is essentially ^{56}Ni (see Fig. 1.14): about $0.7 M_{\odot}$ are produced (about half of the WD mass), as required by the observed luminosity of SNIa. Only in the central regions, the high densities favor the production of ^{54}Fe and ^{58}Ni through electron captures. The overproduction of those nuclei is one of the current shortcomings of this scenario. In the outer layers, NSE is not reached because of lower temperatures, which leads to the production of intermediate mass nuclei (like Si, Ca, etc.). Again, the simplest scenario produces at present too little amounts of those nuclei compared with observations.

Despite those shortcomings, the accretion plus C deflagration scenario seems to be on the right track (see, e.g., Hillebrandt and Niemeyer 2002 and references therein). Key issues currently investigated are the conditions under which the thermonuclear explosion starts (precise density, number and location of ignition points) and the propagation of the nuclear flame through the white dwarf (effects of instabilities, turbulence, initial metallicity, rotation, etc.).

¹⁵ In non-relativistic degenerate gases the pressure scales with density as $P \propto \rho^{5/3}$, while in relativistic degenerate gases $P \propto \rho^{4/3}$.

From the point of view of galactic chemical evolution, SNIa play a very important role, since they produce between 50% and 65% of the Fe-peak nuclei. Their smaller frequency (they are ~ 5 times less frequent than core collapse SN) is compensated by the larger amounts of ^{56}Ni and Fe-peak nuclei produced by their explosion ($0.7 M_{\odot}$ of ^{56}Ni compared to $0.1 M_{\odot}$ on average for core collapse SN).

1.5.4 Production of Intermediate Mass Nuclei (from C to the Fe peak)

In general, it is (and will continue to be) very difficult to test models of stellar nucleosynthesis on a star by star basis. However, we already know that, in a statistical sense, current models are not far from reality. Indeed, adopting current stellar yields (of intermediate and massive stars, as well as of SNIa), an appropriate stellar initial mass function (IMF) and running galactic chemical evolution models for the solar neighborhood, one finds that the abundances of elements between

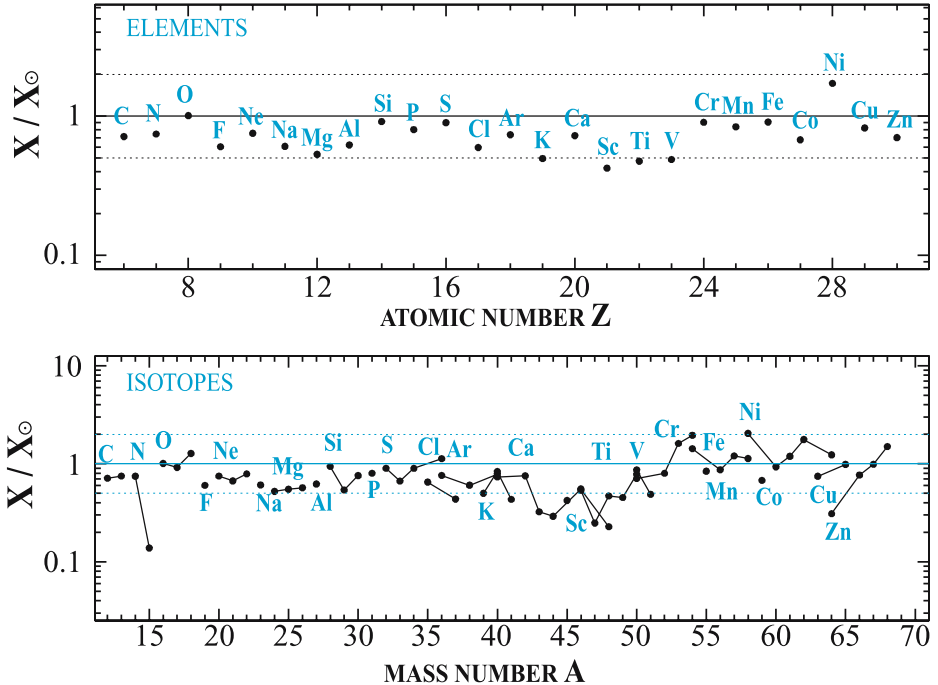


Fig. 1.22. Abundances at solar system formation (4.5 Gyr ago) obtained with a galactic chemical evolution model appropriate for the solar neighborhood and compared to the solar ones. Yields from intermediate and massive stars, as well as SNIa, are used. It can be seen that all elements and isotopes up to the Fe peak are nicely co-produced (within a factor of 2); another source (novae?) is needed for ^{15}N (from Goswami and Prantzos 2000)

C and Zn at solar system formation are satisfactorily reproduced (e.g., Goswami and Prantzos 2000, see Fig. 1.22). This is, at present, the most convincing global test of our theory of stellar nucleosynthesis. Still, observations of abundance ratios in old stars of the Milky Way's halo or in remote gaseous systems (formed early in the history of the universe and enriched predominantly by short-lived massive stars) present new challenges to our understanding of those objects.

1.6 Nuclei Heavier than Fe

Nuclei heavier than those of the Fe peak (henceforth called heavy nuclei) cannot be produced by charged particle reactions. The temperatures needed to overcome the high Coulomb barriers are such that photodisintegrations would tear down all nuclei to their constituent particles. Thus, nature has found another mechanism: neutron captures of pre-existing *seed* nuclei.

1.6.1 Production Mechanisms and Classification of Isotopes

Already in the early works of nucleosynthesis (Burbidge et al. 1957) it was recognized that two different types of neutron capture processes have contributed to the formation of heavy nuclei: the so-called s-process (for slow), responsible for the formation of nuclei in the valley of nuclear stability; and the r-process (for rapid), responsible for the synthesis of the most neutron-rich isotopes of heavy elements.

In the s-process, low neutron densities ($\sim 10^8 \text{ cm}^{-3}$) and rather long timescales ($> 1 \text{ yr}$) are involved, so that when an unstable nucleus (A, Z) is encountered along the s-process path (Fig. 1.23), it has the time to decay into a stable isobar ($A, Z + 1$). Indeed, the lifetimes of most of the unstable nuclei close to the valley of stability range from a few seconds to less than a year and are lower than their lifetimes with respect to neutron captures in the above conditions ($\tau_n = (n_n \sigma_n v)^{-1}$, where n_n is the neutron density, σ_n the neutron capture cross-section and v the average neutron velocity). The neutron captures are slow wrt beta decay and thus, the s-process path closely follows the bottom of the nuclear stability valley. Note that the s-process, by its very nature, cannot reach the heaviest nuclei (the long-lived Th and U isotopes), which are separated from the heaviest stable nucleus (^{209}Bi) by more than 20 mass units.

On the contrary, the r-process involves huge neutron densities ($> 10^{20} \text{ cm}^{-3}$) for timescales of the order of $\sim 1 \text{ s}$ (obviously corresponding to an explosive site): neutron captures are so rapid wrt beta decays that the r-process path drives the stellar material far to the neutron-rich side of the stability valley. When the neutron burst ceases, the unstable nuclei that have been formed decay back to the stability valley. The most neutron-rich nuclei of the heavy elements (pure r-nuclei) are thus formed, as well as the heaviest elements (Th, U, which cannot be reached by the s-process). At the same time, many nuclei that can also be

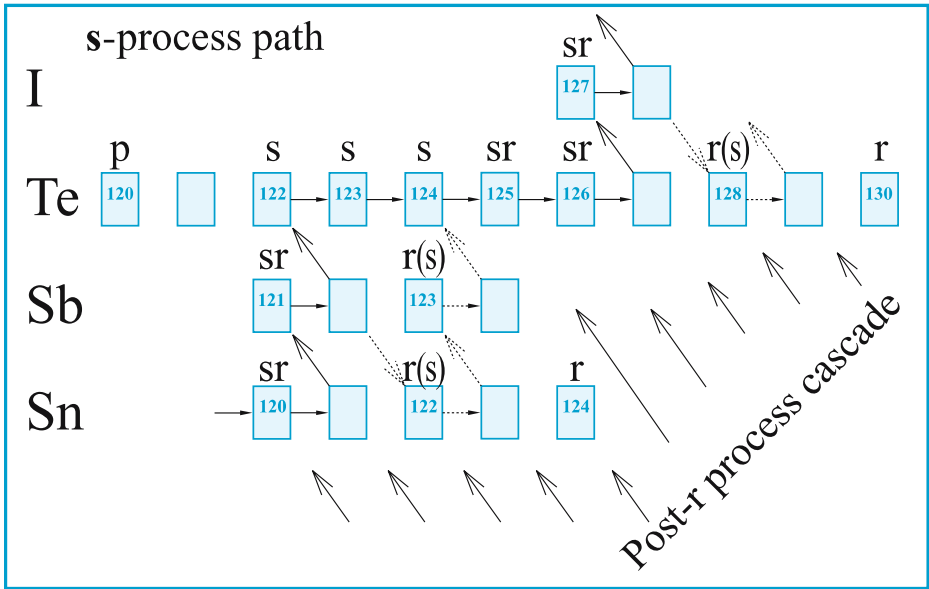


Fig. 1.23. Paths in the plane of neutron number N vs. proton number Z , leading to the formation of nuclei heavier than those of the Fe peak. *Horizontal arrows* indicate neutron captures and *diagonal arrows* beta decays (β^-). For pure (or mostly) s-nuclei the path lies close to the valley of nuclear stability. The most neutron-rich stable nuclei (r-nuclei) are produced by the beta decay of very neutron-rich unstable progenitors, formed by a rapid exposure to high neutron fluxes. The proton-rich p-nuclei cannot be reached by neutron captures (see text)

formed by the s-process are synthesized, so that the abundances of the majority of the heavy nuclei have contributions from both the s- and the r-process, i.e.,

$$N(A, Z) = N_s(A, Z) + N_r(A, Z) . \tag{1.15}$$

Obviously, for pure s-nuclei $N_r(A, Z) = 0$. The abundances of the heavy nuclides, classified as s-, r- or p-nuclei are displayed in Fig. 1.24. The three peaks in the s-nuclei abundance curve correspond to nuclei with *magic* neutron numbers ($N = 50, 82, 126$, see Fig. 1.4). They are accompanied by less sharp peaks in the r-nuclei abundance curve, occurring at smaller mass number A . The total amount of each of the two classes (s- and r-) of heavy nuclide abundances is $\sim 10^{-6}$ by mass; taking into account the very different production mechanisms, this is a rather strange coincidence.

The p-nuclei are much less abundant, by two or three orders of magnitude, than either the corresponding s- or r-isotopes (an exception being ^{92}Mo , which constitutes 14% of Mo). These isotopes are not reached by neutron captures of stable nuclei (Fig. 1.23) and are thought to be produced mainly by photodisinte-

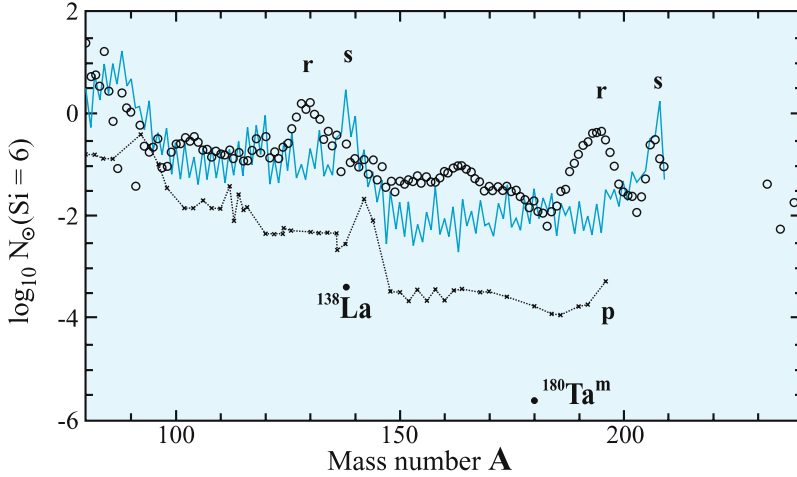


Fig. 1.24. Solar system abundances of heavy nuclei classified as s-, r- and p-nuclei (see Fig. 1.23). ^{180}Ta (probably produced by the p-process) is the rarest isotope in nature

gration of s-nuclei. The most favorable site appears to be the O-Ne layers of core collapse SN explosions, reaching peak temperatures of $2-3 \times 10^9$ K (Arnould and Goriely 2003 and references therein).

1.6.2 The S-Process

An analysis of the solar abundances (Fig. 1.24) and neutron capture cross-sections (Fig. 1.25) of nuclei along the s-process path allows us to determine the relevant physical conditions: neutron densities $n_n \sim 10^8 \text{ cm}^{-3}$, temperatures $T \sim 1.5-3 \times 10^8$ K, densities $\rho \sim 2-12 \times 10^3 \text{ g cm}^{-3}$, and timescales $\tau_S \sim 10-10 \times 10^3$ yr. These conditions correspond to the phase of quiescent He-burning. Two relevant sites have been identified:

1. Central He-burning in massive stars, where the neutron source is $^{22}\text{Ne}(\alpha, n)$. This reaction operates towards the end of He-burning, at $T \sim 2.5 \times 10^8$ K (e.g., Prantzos et al. 1990, The et al. 2000). The number of neutrons released allows to synthesize only nuclei in the mass range $60 < A < 90$ (the so-called weak component of the s-process). Part of those nuclei survive the advanced evolutionary phases of massive stars and are ejected in the interstellar medium by the final supernova explosion. From the theoretical point of view, the He-core s-process is quite robust (despite the uncertainties in the $^{22}\text{Ne}(\alpha, n)$ rate), but it lacks direct observational support.
2. Double-shell (H and He) burning in low mass stars, evolving in the asymptotic giant branch (AGB, running very close and parallel to the red giant branch). During the thermal pulse phase (Fig. 1.26), protons from the H-shell are mixed in regions rich in α and ^{12}C (produced from $3\alpha \rightarrow ^{12}\text{C}$ in the He-shell).

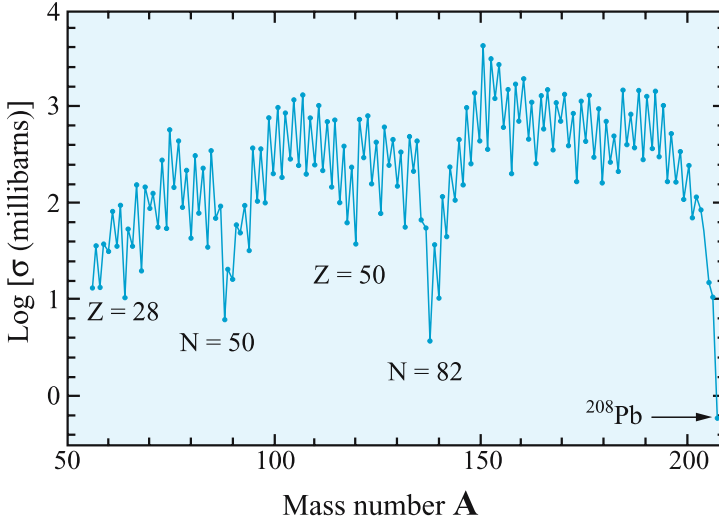


Fig. 1.25. Neutron capture cross-sections of s-nuclei as a function of mass number A . Nuclei with magic neutron numbers (50, 82, 126) have extremely low cross-sections and high abundances with respect to their neighbors (see Fig. 1.24). ^{208}Pb is a representative case, with $Z = 82$ and $N = 126$

The co-existence of p, α and ^{12}C allows the production of abundant neutrons through the reaction series $p + ^{12}\text{C} \rightarrow ^{13}\text{N}$, $^{13}\text{N}(\beta^+)^{13}\text{C}$, $^{13}\text{C}(\alpha, n)^{16}\text{O}$. The recurring pulses allow the production of a large number of neutrons per seed Fe nucleus and the synthesis of the main component of the s-process, i.e., nuclei with $90 < A < 210$. Between thermal pulses the convective stellar envelope penetrates deep in the star and dredges-up material from the s-processed region to the surface, from where the strong AGB wind expels s-nuclei to the interstellar medium. This rather complex scenario is supported by observations of enhanced s-nuclei abundances in the surfaces of low mass AGB stars. However, the mixing of protons with material from the He-zones, as well as the third dredge-up, are poorly understood at present (Busso et al. 1999, Mowlavi and Goriely 2000).

1.6.3 The R-Process

For sufficiently high neutron densities, successive neutron captures convert seed nuclei of a given element to very neutron-rich ones. This continues until nuclei with very low neutron separation energies S_n are encountered (Fig. 1.27). At this point the inverse reactions (γ, n) balance (n, γ) reactions. In the simplest picture of the r-process, a $(n, \gamma) \leftrightarrow (\gamma, n)$ equilibrium is reached in each isotopic chain. Analogously to the NSE (see Sect. 1.4.3) the isotopic abundances of a chain of given Z depend then on S_n , temperature T and neutron density n , as well as

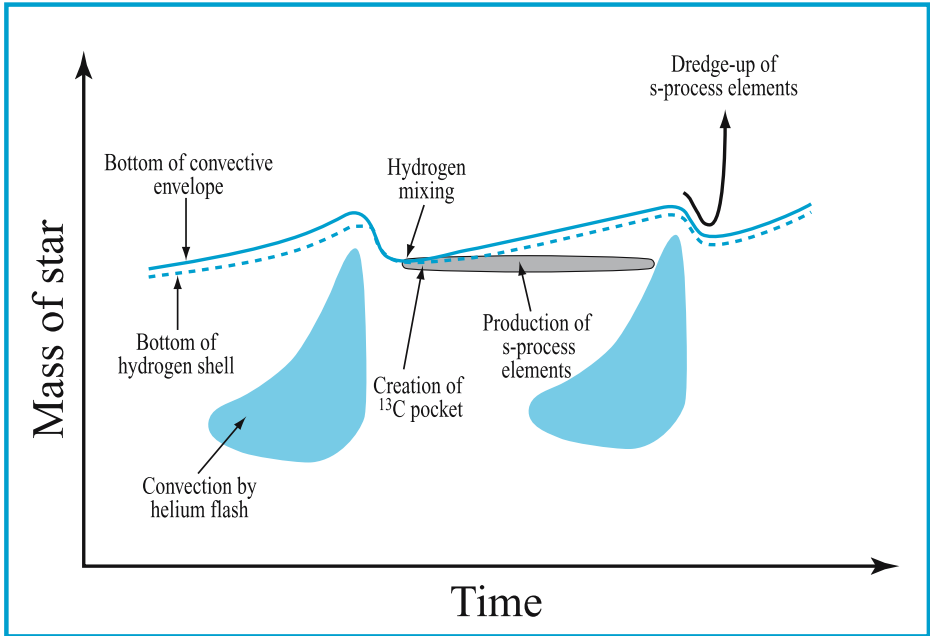


Fig. 1.26. Structure of a thermally pulsing AGB star. After a thermal pulse protons from the H-shell are mixed with He-shell material (rich in ^4He and ^{12}C). This leads to neutron production through $^{12}\text{C}(p,\gamma)^{13}\text{N}(\beta^+)^{13}\text{C}(\alpha,n)$ and s-process nucleosynthesis. After the next thermal pulse, s-nuclei are brought to the stellar surface by the convective envelope and expelled in the interstellar medium

on nuclear partition functions. In those conditions, a large amount of material is accumulated in unstable neutron-rich nuclides with magic neutron numbers. Material slowly shifts to higher Z values through beta decays, which govern the speed of the nuclear flow along the r-process path. At the end of the neutron irradiation, all unstable nuclei beta decay towards the nuclear stability valley. The peaks in the r-nuclei abundance curve (shifted by a few mass units with respect to the corresponding s-nuclei abundance peaks) are obtained from the beta decay of very neutron-rich material “stored” in nuclei with magic neutron numbers during the r-process (e.g., Rolfs and Rodney 1988).

Even for the simplest r-process model, a large body of nuclear data is required (masses and beta decay rates of very unstable neutron-rich nuclei). More sophisticated models drop the assumption of $(n,\gamma)\leftrightarrow(\gamma,n)$ equilibrium and require also neutron capture and photodisintegration rates, along with very large nuclear reaction networks (involving several thousands of nuclei).

The site of the r-process is not yet identified, despite more than thirty years of intense investigation. Required neutron densities and timescales suggest an explosive event, while observations of the composition of the oldest stars in the

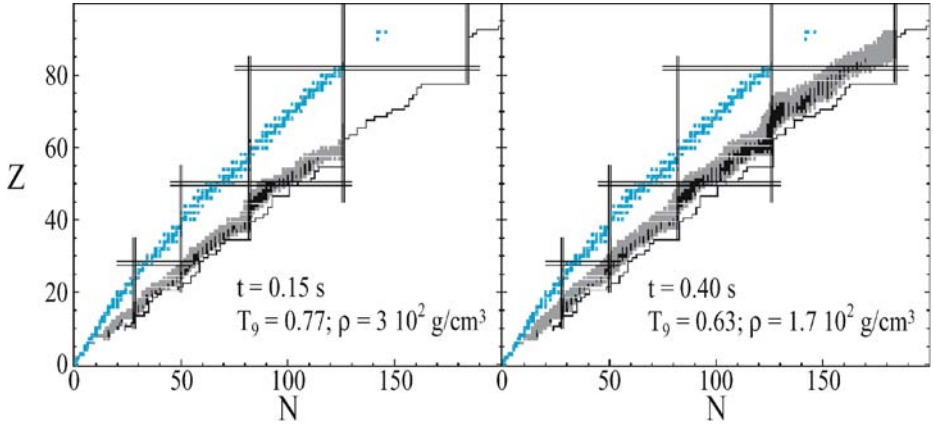


Fig. 1.27. Nuclear flow in a calculation of the dynamical r-process, at two different instances. Neutron captures drive material away from the nuclear stability valley, towards the neutron-drip line (where neutron separation energy $S_n = 0$, curve in the right of the diagrams). *Straight lines* indicate magic nuclei (From Arnould and Takahashi 1999)

Milky Ways suggest that the r-nuclei were produced quite early on, i.e., that their source was short-lived. Core collapse SN satisfy both those requirements, but neither the neutron source nor the relevant stellar layers have been identified yet (e.g., Goriely et al. 2004, Cowan and Thielemann 2004).

1.7 Summary

The basic picture of stellar nucleosynthesis is well established nowadays, at least qualitatively. It can be summarized as follows.

Massive stars ($M > 10 M_{\odot}$) synthesize in their various burning layers elements from C to Ca and eject them in the interstellar medium in the final SN explosion. During the explosion, a large fraction of Fe-peak isotopes (from Ti to Zn) is also produced.

Intermediate mass stars ($2-10 M_{\odot}$) produce important amounts of He and CNO isotopes (except ^{16}O , ^{18}O and ^{15}N), as well as nuclei heavier than Fe-peak nuclei lying on the nuclear stability valley (s-nuclei, produced by neutron captures). All those nuclei are convected to the surface of the star and ejected in the interstellar medium through the stellar winds, mostly during the AGB and planetary nebula phases.

The nuclei heavier than Fe-peak nuclei that are lying relatively away from the nuclear stability valley are thought to be produced during the explosions of massive stars, either in stellar layers submitted to large neutron fluxes (neutron-rich or r-nuclei) or in layers where photodisintegrations of already synthesized s-nuclei play an important role (proton-rich or p-nuclei); the former process

is also responsible for the formation of the heaviest stable nuclei, the Th and U isotopes.

Finally, a large fraction of the Fe-peak nuclei (>50%) is synthesized in thermonuclear SN (SNIa), white dwarfs in binary systems that explode upon reaching the Chandrasekhar limit M_{Ch} (through accretion from a companion star).

Taking into account the lifetimes (a few 10^6 – 10^7 yr for massive stars, a few 10^8 – 10^9 yr for intermediate mass stars and $>10^9$ yr for the average SNIa), as well as the relative numbers of the various nucleosynthesis sites (the so-called stellar initial mass function, which favors the less massive stars), one may construct models of galactic chemical evolution¹⁶ and compare their results to various observables.

The success of this scenario lies mainly in its ability to reproduce (within reasonable assumptions) the solar distribution of intermediate mass elements (C to Fe peak, see Fig. 1.22) and s-elements. It also explains reasonably well most of the abundance ratios of such elements that are observed on the surfaces of stars of various ages in the Milky Way.¹⁷

Several open questions however remain (e.g., the proton mixing at the origin of the $^{13}\text{C}(\alpha, n)$ neutron source for the s-process in AGB stars, the stellar layers of the core-collapse SN where the r-process takes place, etc.). Furthermore, uncertainties in key nuclear data (e.g., the $^{12}\text{C}(\alpha, \gamma)$ rate for He-burning, the nuclear masses for the r-process) or in important stellar physics processes (e.g., mixing processes in the stellar core and envelope, the explosion mechanism in core collapse SN, the role of rotation, mass loss and even magnetic fields in overall stellar evolution, etc.) prevent us from a satisfactory quantitative picture at present. Progress is expected from continuous improvements in the fields of numerical simulations, nuclear data and spectroscopic observations in all wavelengths.

The theory of stellar nucleosynthesis is certainly important to exobiology studies, to the extent that life (at least as we currently understand it) requires the presence of various heavy elements (e.g., oxygen, carbon, etc.). However, a quantitative assessment of that requirement is difficult at present. Indeed, the absolute amount of heavy elements (i.e., their abundance with respect to H) that is required for the formation of telluric planets is very poorly constrained, either from theory or from observations. The only relative (but not necessarily relevant) constraint at present is the metallicity distribution of stars with giant gaseous planets around them. Observations in recent years (e.g., Santos et al. 2003) suggest that such stars are more Fe-rich on average than stars of the same

¹⁶ Chemical evolution in this context means the evolution of the abundances of the various chemical elements.

¹⁷ Some important failures of our current understanding of stellar nucleosynthesis to interpret observations of abundance ratios in the most metal poor stars of our galaxy, may point to significant variations from the “standard picture”, at least as far as the very first generations of massive stars are concerned (they could be more massive, on average, than today’s stars, or have larger rotation velocities; see, e.g., Prantzos 2004 and references therein).

spectral type not hosting planets around them. In particular, almost all planet hosting stars have a metallicity larger than 1/3 of the corresponding solar one (while disk stars in the solar neighborhood have metallicities extending down to at least 1/10 of solar and even less).

This observational fact does not have a sound theoretical understanding at present, and theoreticians of planetary system formation are only now starting to considering such constraints for their models (e.g., Pinotti et al. 2005). Assuming that this metallicity constraint applies also to the formation of telluric planets, one concludes that such planets should exist fairly early in the history of the local galaxy, since its metallicity reached 1/3 of solar metallicity several gigayears before the formation of our solar system. The formation of such planets would be even more favored in the inner galaxy, where metallicities are higher than in the solar neighborhood; the opposite should apply to the outer disk of the Milky Way or to the nearby galaxy of the Large Magellanic Cloud, where the maximum metallicity today is barely 1/3 that of solar metallicity. On the contrary, there would be no chance to find such planets in the metal poor stars of the galactic halo or of the Small Magellanic Cloud, since their metallicities never exceed 1/10 of the solar metallicity.

In summary, there is certainly a relationship between the history of the abundances of chemical elements in the universe and the formation of telluric planets (and life). That relationship is very poorly understood at present, but it is expected that observations of extrasolar telluric planets with future instruments (e.g., ESA's DARWIN satellite) will offer us some hints.

References

General References

- Audouze J., Vauclair S. (1995) *L'Astrophysique Nucléaire*. Presses Universitaires de France, Paris
- Cowan J., Thielemann K.-F. (2004) *Physics Today*, October, p. 47
- Cowley C. (1995) *An Introduction to Cosmochemistry*. Cambridge University Press, Cambridge
- Pagel B. (1997) *Nucleosynthesis and Galactic Chemical Evolution*. Cambridge University Press, Cambridge
- Prantzos N., Montmerle T. (1998) *Naissance, Vie et Mort des Etoiles*. Presses Universitaires de France, Paris

Specialized References

- Arnett D. (1996) *Supernovae and Nucleosynthesis*. Princeton University Press, Princeton
- Arnett D., Bahcall J., Kirshner R., Woosley S. (1989) *ARAA* 27, 629
- Arnould M., Takahashi K. (1999) *Rep Prog Phys* 62, 395
- Arnould M., Goriely S. (2003) *Phys Rep* 384, 1

- Asplund M., Grevesse N., Sauval J. (2005) In: Bash F.N., Barnes T.G. (eds.) *Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis*, ASP Conference Series, vol. 336. San Francisco
- Bazan G., Arnett D. (1994) *ApJL* 433, 41
- Burbidge M., Burbidge G., Fowler W., Hoyle F. (1957) *Rev Mod Phys* 29, 547
- Busso M., Gallino R., Wasserburg G. (1999) *ARAA* 37, 239
- Cameron A.G.W. (1957) *PASP* 69, 201
- Chieffi A., Limongi M., Straniero O. (1998) *ApJ* 502, 737
- Colgate S., Grasberger W., White R. (1961) *AJ* 70, 280
- Goriely S., Demetriou V., Janka H.-Th., Pearson M., Samyn M. (2004) In: Buchmann L. et al. (eds.) *Nuclei in the Cosmos VIII*, Vancouver, 19–23 July 2004
- Goswami A., Prantzos N. (2000) *AA* 359, 191
- Guidry M. (1998) In: J. Hirsch, D. Page (eds.) *Nuclear and Particle Astrophysics*, p. 115. Cambridge University Press, Cambridge
- Hamuy M. (2003) *ApJ* 582, 905
- Hartmann D., Woosley S., El Eid M. (1985) *ApJ* 297, 837
- Heger A., Fryer C., Woosley S., Langer N., Hartmann D. (2003) *ApJ* 591, 298
- Hillebrandt W., Niemeyer J. (2002) *ARAA* 38, 191
- Janka H.-T. (2001) *AA* 368, 527
- Janka H.-T., Buras R., Kifonidis K., Rampp M., Plewa T. (2002) In: Fryer C.L. (ed.) *Core Collapse of Massive Stars*. Kluwer, Dordrecht
- Limongi M., Straniero O., Chieffi A. (2000) *ApJS* 129, 625
- Lodders K. (2003) *ApJ* 591, 1220
- Maeder A., Meynet G. (2003) *AA* 404, 975
- Maeder A., Meynet G. (2002) *ARAA* 38, 143
- Mowlavi N., Goriely S. (2000) *AA* 362, 599
- Pinotti R., Arany-Prado L., Lyra W., Porto de Mello G. (2005) *MNRAS* 364, 29
- Prantzos N. (2000) *New Astr Rev* 44, 303
- Prantzos N. (2004) In: Buchmann L. et al. (eds.) *Nuclei in the Cosmos VIII*, Vancouver, 19–23 July 2004
- Prantzos N., Hashimoto M., Nomoto K. (1990) *AA* 234, 211
- Rols C., Rodney R. (1988) *Cauldrons in the Cosmos*. Chicago University Press, Chicago
- Santos N., Israelian G., Mayor M., Rebolo R., Udry S. (2003) *AA* 398, 363
- The L.S., El Eid M., Meyer B. (2000) *ApJ* 533, 998
- Thielemann F.-K., Arnett D. (1985) *ApJ* 295, 604
- Thielemann F.-K., Nomoto K., Hashimoto M. (1996) *ApJ* 460, 408
- Thielemann F.-K., Rauscher T., Freighburghaus C. et al. (1998) In: J. Hirsch, D. Page (eds.) *Nuclear and Particle Astrophysics*, p. 27. Cambridge University Press, Cambridge
- Woosley S. (1986) In: Hauck B., Maeder A., Meynet G. (eds.) *Nucleosynthesis and Chemical Evolution*, Proceedings of the 16th Advanced Course of the SSAA, Geneva, 17–27 March 1986
- Woosley S., Weaver T. (1995) *ApJS* 101, 181
- Woosley S., Arnett D., Clayton D. (1973) *ApJS*, 26, 231
- Woosley S., Heger A., Weaver T. (2002) *Rev Mod Phys* 74, 1015

2 Formation of the Solar System: a Chronology Based on Meteorites

Marc Chaussidon

2.1 In Search for Ages

Knowing the age of the solar system or that of the Universe, the Earth or the first appearance of man, has always been a fundamental question. On top of the philosophical aspect of this search for ages, there is a better, more *down to Earth* reason to estimate these ages with a continuously improved accuracy. This reason is to understand the processes involved during the formation of the solar system. Finding out when and how fast the Sun and the planets formed (in 10^5 , 10^6 , or 10^7 years) allows to select between different physical processes and to compare our system with the young stars and/or the extrasolar planetary systems that the astrophysicists are now observing with more and more precision and detail.

Two complementary approaches provide information on the chronology of the solar system's formation: (1) the dating of samples illustrating the various evolutionary steps of the solar system materials (meteorites and other extraterrestrial samples, which are dated by geochemical methods using their radioactivity), or (2) the physical and chemical modeling of the processes involved (e.g., how much time is needed for the Sun to accrete? How long does the formation of a planetesimal or a giant planet take, etc.).

It would be erroneous to believe that the geochemical ages obtained on meteorites are always *accurate and precise* and that they do not depend on models as the ages obtained by modeling the formation of the solar system. In fact, despite their perceived precision, some geochemical ages are also based on regularly questioned hypotheses.

Several reviews concerning the chronology of the solar system and the differentiation of the planetesimals, the origin of extinct radioactivities or the origin of isotopic variations observed in meteorites have recently been published (Wadhwa and Russell 2000; Goswami and Vanhala 2000; Russell et al. 2001; McKeegan and Davis 2003; Birck 2004; Chaussidon and Gounelle 2006). Older articles are also relevant (e.g., Wetherill 1975; Wasserburg 1987; Podosek and Swindle 1988a, b; Tilton 1988a, b). These articles offer a detailed summary of the existing literature on the subject. The purpose of this chapter is not to duplicate these reviews but rather to present a few examples of geochemical ages fundamental in our understanding of the formation of the solar system, to evaluate them in

a critical way, to estimate their relative robustness and to emphasize important but unanswered questions.

2.2 What is a Geochemical Age?

Geochemical ages rely on the natural radioactivity of rocks. In short, they are obtained by measuring the quantity of daughter isotopes produced in a sample from the radioactive decay of its parent isotope. Knowing the half-life of the radioactive decay, the time required to accumulate such a quantity of daughter isotopes can be calculated.

2.2.1 The Radioactivity

The natural radioactivity of rocks is due to the presence of radioactive isotopes. A large number of radioactive isotopes exists in nature and some of them are currently used to date rocks (see Faure 1986; Vidal 1998; Albarède 2003; Allègre 2005; for a detailed description of the principles of isotopic dating in geology and for some basic applications). Several types of radioactive decay exist (see for example Faure 1986) but since radioactivity is a nuclear process, the rate of decay is constant and thus independent of all the physical or chemical processes that a rock may have undergone. Therefore, it provides a reliable clock whose utilization may nevertheless be sometimes problematic or even impossible. The rate of decay (noted λ) is defined as the probability that a radioactive element decays per unit of time

$$\lambda = -dN/(N \times dt) . \quad (2.1)$$

By integration, the law of radioactive decay is obtained

$$N = N_0 \times e^{-(\lambda \times t)} \quad (2.2)$$

where N is the number of atoms of radioactive isotope remaining at time t and N_0 the number of atoms initially present at time $t = 0$. The decay constant (λ) is characteristic of the element considered and has a t^{-1} dimension. To simplify, the notion of half-life is commonly used in place of λ . The half-life ($T_{1/2}$) is the time required for the decay of half of the number of radioactive atoms originally present

$$T_{1/2} = (\ln 2)/\lambda . \quad (2.3)$$

In Anglo-Saxon literature, the notion of the mean life (τ) is sometimes used: τ is the inverse of the decay constant ($\tau = 1/\lambda$).

Usually, the long period radioactive elements are distinguished from the short period radioactive elements, which, if present in the solar system 4.5 billion years

Table 2.1. List of short-lived nuclides (isotopic ratio of radioactive isotope over stable isotope) that can be used to establish a relative chronology for the early solar system. The initial ratios are from the references indicated in last column. The exact values of an initial ratio remain a matter of debate, as well as the concept of an initial ratios for some of these systems (see text).

Isotopic ratio	Half-life (Ma)	Initial ratio	Reference
${}^7\text{Be}/{}^9\text{Be}$	53 days	6×10^{-3}	Chaussidon et al. 2006
${}^{41}\text{Ca}/{}^{40}\text{Ca}$	0.10 Ma	1.5×10^{-8}	Srinivasan et al. 1994
${}^{26}\text{Al}/{}^{27}\text{Al}$	0.7 Ma	5×10^{-5}	Lee et al. 1976
${}^{60}\text{Fe}/{}^{56}\text{Fe}$	1.5 Ma	$<1.6 \times 10^{-6}$	Birck and Lugmair 1988; Tachibana and Huss 2003; Mostefaoui et al. 2004
${}^{10}\text{Be}/{}^9\text{Be}$	1.5 Ma	9.5×10^{-4}	McKeegan et al. 2000
${}^{53}\text{Mn}/{}^{55}\text{Mn}$	3.7 Ma	3.7×10^{-5}	Birck and Allègre 1985
${}^{107}\text{Pd}/{}^{108}\text{Pd}$	6.5 Ma	2×10^{-5}	Kelly and Wasserburg 1978
${}^{182}\text{Hf}/{}^{180}\text{Hf}$	9 Ma	1.6×10^{-4}	Quitté and Birck 2004
${}^{129}\text{I}/{}^{127}\text{I}$	15.9 Ma	1×10^{-4}	Jeffery and Reynolds 1961

ago (Ga), may be considered as extinct today, considering the precision of analytical techniques (if $T_{1/2} = 150\text{Ma}$, the ratio N/N_0 4.5 Ga after the formation of the solar system has a value of $\sim 10^{-9}$). They are thus referred to as extinct radioactivities. These two types of radioactive elements may be used to obtain time information. The long period radioactive elements are used for dating, that is, to measure absolute ages. The extinct radioactivities give access to durations, without providing absolute timescales. The list of the radioactive elements most commonly used in cosmochemistry is given in Table 2.1.

2.2.2 The Absolute Ages

The absolute age is obtained from the measurement in the sample to be dated of the quantities of daughter and parent isotopes. If the system remained close since the formation of the sample, each parent isotope (noted P) has decayed to produce a daughter isotope (noted D) and the mass balance can be written as

$$P_0 - P = D - D_0 . \quad (2.4)$$

In this equation, P_0 and D_0 are unknowns that cannot be measured in the sample today. However, one of them (P_0) can be written, using the law of radioactivity, as a function of what can be measured today. This gives

$$D = D_0 + P \times (e^{(\lambda \times t)} - 1) . \quad (2.5)$$

This equation is then transformed, by dividing each member by the number of atoms of a stable isotope (neither radioactive nor radiogenic) of the daughter element. Isotopic and elemental ratios are in fact the easiest quantities accessible to mass spectrometry. This leads to what is called an isochron equation written, for example, in the case of the $^{87}\text{Rb}/^{87}\text{Sr}$ system (^{87}Rb β^- decays to ^{87}Sr with a half-life of 48.81 Ga) as

$$\left(^{87}\text{Sr}/^{86}\text{Sr}\right) = \left(^{87}\text{Sr}/^{86}\text{Sr}\right)_0 + \left(^{87}\text{Rb}/^{86}\text{Sr}\right) \times \left(e^{(\lambda \times t)} - 1\right). \quad (2.6)$$

This isochron equation can be graphically solved if, in the sample to be dated, several fractions having different parent/daughter ratios ($^{87}\text{Rb}/^{86}\text{Sr}$) can be analyzed. In such a case, and if the system remained closed since its formation, the daughter isotopic ratios plot along a line (isochron) as a function of the parent/daughter elementary ratios. The slope ($e^{(\lambda \times t)} - 1$) gives access to the age and the zero-intercept to the isotopic composition of the daughter element when the sample formed.

2.2.3 The Relative Ages

Considering an extinct radioactivity today, the approximation can be made that in a sample formed ~ 4.5 Ga ago, the concentration of the remaining parent nuclide (P) is equal to zero. In this case, the mass balance equation is simplified and is written as

$$P_0 = D - D_0 \quad \text{or} \quad D = D_0 + P_0. \quad (2.7)$$

For the same reasons as above, this equation can be transformed by using isotopic ratios, for example, as in the case of the $^{26}\text{Al}/^{26}\text{Mg}$ system (^{26}Al β^- decays to ^{26}Mg with a half-life of 0.7 Ma)

$$\left(^{26}\text{Mg}/^{24}\text{Mg}\right) = \left(^{26}\text{Mg}/^{24}\text{Mg}\right)_0 + \left(^{26}\text{Al}/^{24}\text{Mg}\right). \quad (2.8)$$

Time does not appear directly in this equation but is present through the isotopic ratio of the parent nuclide according to

$$\left(^{26}\text{Mg}/^{24}\text{Mg}\right) = \left(^{26}\text{Mg}/^{24}\text{Mg}\right)_0 + \left(^{26}\text{Al}/^{27}\text{Al}\right)_0 \times \left(^{27}\text{Al}/^{24}\text{Mg}\right). \quad (2.9)$$

This isochron equation can also be graphically solved if, in the same sample, several fractions having different parent/daughter ratios ($^{27}\text{Al}/^{24}\text{Mg}$) can be analyzed. In this case, and again if the system remained closed since its formation, once more the daughter isotopic ratios plot along a line (isochron) as a function of the parent/daughter elemental ratios. The slope $\left(^{26}\text{Al}/^{27}\text{Al}\right)_0$ and the zero-intercept $\left(^{26}\text{Mg}/^{24}\text{Mg}\right)_0$, respectively, give the isotopic composition of the parent and daughter elements when the sample was formed. Since the considered parent isotope is a short-period radioactive nuclide, its isotopic composition is changing rapidly with time (Fig. 2.1). The $^{26}\text{Al}/^{27}\text{Al}$ ratio decreases, for instance, by

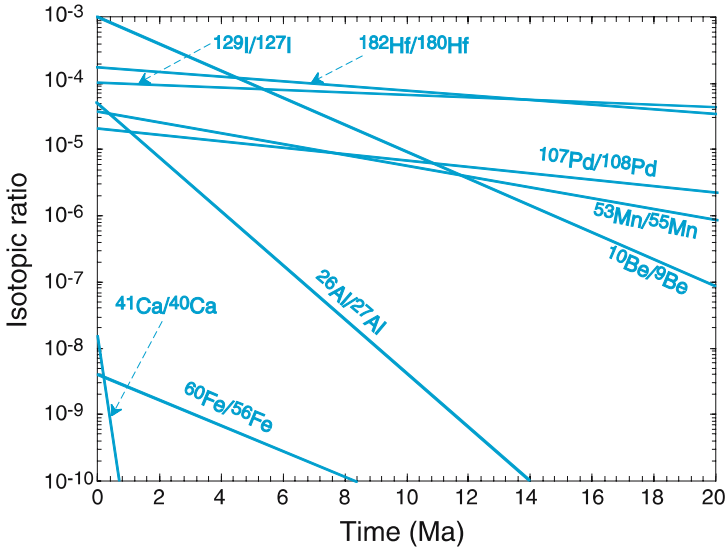


Fig. 2.1. Variations of the isotopic ratios of extinct radioactivities as a function of time (see Table 2.1 for half-lives and initial ratios used). The abundance of extinct radioactivities decreases in the solar system as soon as the nucleosynthetic processes stop, theoretically defining a time 0. One problem is that this time is most probably not the same for all extinct radioactivities because some of them are pre-solar when others (in particular ${}^7\text{Be}$ and ${}^{10}\text{Be}$) are produced in the forming solar system by the irradiation of a part of the accretion disk by the young Sun. In addition, recent Ni isotopic analyses of CAIs, chondrites and iron meteorites suggest that ${}^{60}\text{Fe}$ of supernova origin might have been injected in the solar nebula $\approx 600\,000$ years after CAI formation (Bizzarro et al. 2006). Otherwise, it is easy to understand why the analysis of various extinct radioactivities of the same sample should allow for its formation age to be obtained very precisely

a factor of 2 in 0.7Ma. Thus, two samples 1 and 2 formed from the same original reservoir at different times will show a formation age difference $\Delta t = t_1 - t_2$ that can be written as a function of the isotopic ratios of the parent nuclides according to, in the case of ${}^{26}\text{Al}$ for instance,

$$\left({}^{26}\text{Al}/{}^{27}\text{Al}\right)_{t_1} / \left({}^{26}\text{Al}/{}^{27}\text{Al}\right)_{t_2} = e^{-\lambda \times \Delta t} . \quad (2.10)$$

In theory, variations in short-lived radioactive nuclides isotopic compositions should allow to build a relative chronology between several samples formed from the same reservoir.

2.2.4 Sources of Error or Uncertainty in Isotopic Dating

Obviously, several conditions have to be met so that an absolute or relative age can be obtained based on the two approaches previously described. The samples

have of course to be cogenetic, but the most important condition is that the rock or the mineral to be dated has remained in a closed system for both the parent and daughter isotopes since its formation. These restrictions apply to any geological system but a number of other problems, specific to the earliest evolution of the solar system, must be mentioned.

If a mineral condenses from a gas at high temperature or crystallizes from a magmatic liquid, the system remains open as long as the isotopic compositions have time to homogenize themselves by diffusion in the liquid or the solid. This homogenization is one of the necessary conditions for dating: isotopic homogeneity must have been achieved in the parental magmatic liquid for the different phases crystallized out of this liquid to share the same initial isotopic composition and to line up on the same isochron. In general, this condition is fulfilled in the case of terrestrial magmatic systems that crystallize over rather long timescales ($>10^5$ or 10^6 years), although in some very specific cases the melting processes may produce isotopically heterogeneous liquids (Hammouda et al. 1996). In the case of meteorites, because melting processes could have been very short, i.e., a few minutes typically for the flash-melting events proposed to be at the origin of chondrules (e.g., Hewins 1997), it is conceivable that isotopic homogeneity may not have been reached for all elements.

During the cooling of a mineral, an isotopic closing temperature can be defined as corresponding to the temperature at which the diffusion is too slow to homogenize the isotopic compositions at the scale of a hundred μm or mm. The closing temperatures depend on the diffusion rate of the elements considered, which may vary by orders of magnitude between different elements (e.g., two orders of magnitude between Li and the other elements, Richter et al. 2003). The cooling rates deduced from experiments are very rapid for objects formed in the solar nebula, between ~ 0.1 to $\sim 10\text{K/hour}$ for the Ca and Al-rich refractory inclusions or CAIs (Stolper and Paque 1986), and between a few to maybe a few thousands degrees per hour for the chondrules (Hewins 1997). However, this is not at all the case for the magmatic processes taking place in protoplanetary bodies, a few kilometers to a few hundreds kilometers in size. Thus, if a crystallization age in the nebula can be derived for a CAI or a chondrule, this is not always straightforward for the components of differentiated meteorites (or even of chondrites, primitive meteorites) whose parent bodies, without having been melted, went through a complex metamorphic history after accretion. The study at high resolution of the U/Pb system in phosphates from ordinary chondrites has shown, for instance, that the closing temperature of these minerals was achieved over a period that could reach 60Ma (Göpel et al. 1994). This interval is the time necessary for the post-accretion cooling of the parent body. It depends on several factors such as the depth within the parent body, its size, and its composition. This period is in fact not connected to the duration of the accretion of the parent bodies of ordinary chondrites but represents the time necessary to dissipate the heat produced by the accretion and the decay of short-period radioactive elements (e.g., ^{26}Al or ^{60}Fe).

Apart from all the problems related to the complex history of meteorites (e.g., heterogeneity of the nebula, thermal metamorphism in the parent bodies, shock metamorphism during asteroidal collisions, terrestrial alteration, etc.) other limitations to the determination of an isotopic age must be considered. First, the precision of isotopic analyses may be limited by the corrections due to blanks or the poor statistics due to the limited amount of material available for analysis. In the case of U/Pb system (^{235}U decays to the stable isotope ^{207}Pb with a half-life of 0.7 Ga, ^{238}U to ^{206}Pb with a half-life of 4.47 Ga see Faure 1986 for a detailed description) for instance, this may restrict the precision of the age to a range from around 0.5 to a couple of Ma (e.g., Göpel et al. 1994; Amelin et al. 2002). In some specific cases, an error may derive from uncertainties in the value of the decay constant. This effect is amplified for very old ages, as is the case for meteorites. A symptomatic example is the $^{176}\text{Lu}/^{176}\text{Hf}$ system (^{176}Lu β^- decays to ^{176}Hf with a half-life of ≈ 35.7 Ga). The measurements of the decay constant of ^{176}Hf show a $\sim 4\%$ variability (Begemann et al. 2001), which should lead to an error of about 15 Ma on the Lu/Hf age obtained on chondrites. Recent works (Blichert-Toft et al. 2002; Patchett et al. 2004) aim to improve the value of this decay constant by analyzing meteorites of known ages. However, using the decay constant determined from terrestrial rocks of very precisely known ages result in a 4% “too old” age for chondrites (Amelin and Davis 2005). This apparent discrepancy between ^{176}Lu - ^{176}Hf ages of chondrites and of terrestrial rocks may be solved if the decay rate of ^{176}Lu was “accelerated” in the early solar system by gamma-ray irradiation that could transform a fraction of ^{176}Lu in the ground state to one of its isomer which has a very short half life of 3.7 hours (Albarède et al. 2005).

In conclusion, the significance of an isotopic age relies on a number of hypotheses, which cannot always be rigorously or independently tested. The reconstruction of the early solar system chronology is thus based on comparing the results of different isotopic systems, the different systems not always being equivalent in terms of *quality* of age reconstruction.

2.3 What are the Processes that can be Dated by Isotopic Analyses of Meteorites?

Meteorites are classified into different families or groups that correspond to rocks formed either in different regions of the solar nebula or at different stages of its evolution or both. Primitive meteorites, or chondrites, represent fragments of small size planetesimals that did not undergo major differentiation. For example, the Orgueil carbonaceous chondrite has the same chemical composition as the Sun (except for gases of course) showing that it results from the condensation, without significant chemical fractionation, of mineral phases from the gas of the nebula. In contrast, the differentiated meteorites (achondrites, iron meteorites or iron-silicate meteorites) are fragments of larger planetesimals or planets (Mars,

the Moon), which have gone through several stages of differentiation. By dating various components of these meteorites, a chronology of the formation and the early evolution of the solar system can be established.

Primitive chondrites contain three major components: refractory inclusions rich in Ca and Al (or CAIs from “Ca-Al-rich inclusions”), chondrules, and the matrix embedding the chondrules. CAIs and chondrules are considered to have formed in the nebula during high temperature processes. The matrix contains minerals stable at lower temperatures, minerals that might have condensed in the nebula but that were most likely also affected by hydrothermal or metamorphic processes on parent bodies. Because of their refractory composition, the CAIs are regarded as samples of the first solids condensed from the gas in the nebula at high temperature ($>1800\text{K}$). The chondrules are less refractory. The processes of CAI and chondrule formation are still poorly understood but they likely included one or several of the following processes: (1) condensation from the nebular gas (to form solid or liquid precursor phases), (2) accretion of grains of various origins either condensed from the gas or fragmented from previously existing objects, (3) melting and/or crystallization of the accreted solids and liquids, and (4) reactions between gas and silicate liquids. Multiple episodes of melting, crystallization and alteration are frequently considered to explain the mineralogical, chemical and isotopic diversity of the different components of chondrites.

A geochemical fractionation between the parent and daughter nuclides is required to obtain an isotopic age. The dating of the accretion processes leading to the formation of chondrites seems for instance to be out of reach of the isotopic clocks because no major chemical redistribution or fractionation took place since accretion occurred at low temperatures (“cold” accumulation of chondrules, CAIs and the matrix). Three major processes induced chemical fractionation during the early evolution of the solar system: condensation, melting/crystallization, and metal silicate-sulfide differentiation.

- *Condensation* induces a fractionation between gas and solid phases. Its intensity depends on the condensation’s temperature and the geochemical behavior of the considered elements, or in other words, of their refractory or volatile character (forming solids stable either at high temperatures or at low temperatures, respectively).
- *Melting/crystallization* (and metamorphism) induce fractionations between liquids (silicate or metal) and solids (minerals). Their intensities depend on the condition of the reactions (crystallization rate, temperature) and the geochemical behavior of the considered elements, which are either incompatible (being soluble in liquids) or compatible (being soluble in the minerals that are stable in contact with these liquids).
- *Metal-silicate-sulfide differentiation* induces fractionations between silicate, metal and sulfide liquids. Their intensity depends on the lithophile (having a strong solubility in the silicate liquids), siderophile (having a strong solubility in the metal liquids) or chalcophile (having a strong solubility in the sulfide liquids) character of the elements.

2.4 From the First Solids to the First Planets: When and How Fast?

2.4.1 The Age of Meteorites and the Duration of Accretion Processes: the First Order Picture

Meteorites are the oldest rocks known in the solar system. They have an absolute age of about 4.55 Ga (Wasserburg 1987). A precise analysis of the U/Pb system in CAIs has shown that they are the oldest component of chondrites: 4566^{+2}_{-1} Ma (Manhès et al. 1988; Allègre et al. 1995). This age was recently refined to 4567.2 ± 0.6 Ma (Amelin et al. 2002). It is logical to consider that this is the oldest measurable age in the solar system, since in meteorites the CAIs are the phases condensed at the highest temperature, thus likely the earliest in the history of the solar system. We shall see in the following that some hints exist for slightly older ages.

The $^{87}\text{Rb}/^{87}\text{Sr}$ system provided a first indication that nebular processes lasted several millions years. The eucrites (a group of differentiated meteorites) define an isochron whose slope gives an age of 4.47 ± 0.24 Ga and an initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of 0.698990 ± 0.000047 . The latter represents the Sr isotopic ratio of the eucrites parent body at the time of its formation. It is higher (by $0.31 \pm 0.03\%$) than the one measured in some refractory inclusions of the Allende chondrite. For a solar Rb/Sr (parent/daughter) concentration ratio, this 0.31% difference in the initial $^{87}\text{Sr}/^{86}\text{Sr}$ suggests a duration of ~ 9 Ma of ^{87}Rb decay (if a chondritic Rb/Sr ratio is used, the 0.31% represents a duration of ~ 14 Ma). A first order interpretation of these data is that the nebula lasted some 10 Ma. The oldest samples available are the refractory inclusions, which condensed during a short time interval from the nebular gas at very high temperatures. The mineral precursors of the eucrites parent body likely condensed 10 Ma later from an already colder nebula.

2.4.2 A Relative Chronology Based on the Extinct Radioactivity of ^{26}Al

For the past thirty years, studies have shown that one of the major causes of the variation in meteorites of the Mg isotopic ratio ($^{26}\text{Mg}/^{24}\text{Mg}$) was the in situ radioactive decay of ^{26}Al (^{26}Al β^- decays to ^{26}Mg with a half-life of 0.7 Ma). Since the discovery that “live ^{26}Al ” was incorporated in CAIs of the CV3 Allende chondrite during their formation (Lee et al. 1976), the presence of a ^{26}Mg excess due to the in situ decay of ^{26}Al (isochron relationship in a $^{26}\text{Mg}/^{24}\text{Mg}$ vs. $^{27}\text{Al}/^{24}\text{Mg}$ diagram) was demonstrated in chondrules, chondrites, and the differentiated meteorites. Recently, technical improvements allowed to obtain very precise Mg isotopic analysis for bulk samples (Bizzarro et al. 2004, 2005b; Baker et al. 2005). The variations of the initial $^{26}\text{Al}/^{27}\text{Al}$ ratios (slopes of the isochron diagrams, see Fig. 2.2) of these different objects allow to build

a relative chronology from CAIs to the differentiated meteorites by making the hypothesis of an homogeneous initial distribution of ^{26}Al in the solar accretion disc. Given this assumption, the highest $^{26}\text{Al}/^{27}\text{Al}$ ratio corresponds to the oldest objects (or events) and the decrease of the $^{26}\text{Al}/^{27}\text{Al}$ ratio can be translated in time elapsed using the law of radioactivity. This assumption is made in the next section to investigate its consequences in terms of relative chronology, but it is important to note that this remains an open question, which will be discussed further in Sect. 2.5.2.

The $^{26}\text{Al}/^{27}\text{Al}$ ratio of CAIs at the time of their formation (derived from the excess in ^{26}Mg measured) ranges from $\sim 5 \times 10^{-5}$ to 0. Most of the CAIs have a ratio close to 5×10^{-5} (mode of analyses 4.5×10^{-5}), which has been classically referred to as the canonical ratio (see synthesis of MacPherson et al. 1995; Swindle et al. 1996; Wadhwa and Russell 2000) despite CAIs do show a range in $^{26}\text{Al}/^{27}\text{Al}$. The Al/Mg fractionation observed in the ^{26}Al isochrons of CAIs occurs between the various CAI mineral phases during the fractional crystallization of their parent liquid in the gas of the nebula. By assuming that this crystallization is rapid (e.g., Stolper and Paque 1986), it can be considered that the crystallization age of the CAIs is very close to that of the high-temperature condensation processes, which have generated the precursor grains or liquids. However, this may not be always the case because the petrology of CAIs shows that a succession of several melting events could have occurred. The CAIs having a $^{26}\text{Al}/^{27}\text{Al}$ ratio close to 5×10^{-5} (Fig. 2.2a) are generally the ones for which the $^{26}\text{Al}/^{27}\text{Al}$ ratio is best determined (they have high $^{27}\text{Al}/^{24}\text{Mg}$ and/or little alteration for example). Many secondary processes may in fact disturb the distribution of Mg and Al in CAIs and transform an original isochron with a 5×10^{-5} slope into a *wrong isochron* with lower or higher slopes (Podosek et al. 1991). Still considering the hypothesis of a homogenous $^{26}\text{Al}/^{27}\text{Al}$ ratio for the initial nebula, the variation range (from 5×10^{-5} to 0) may be interpreted in two different ways. This range reflects the duration of the high-temperature condensation processes in the nebula, and in this case, they would have lasted around 5 Ma. Alternatively, this variation is an artifact caused by magmatic perturbations and the high-temperature condensation processes could be a very short event lasting about one Ma or less. Finally, recent analyses have opened a debate on the initial $^{26}\text{Al}/^{27}\text{Al}$ value of the CAIs. Based on bulk rock CAI measurements, Galy et al. (2004) have shown that some of them define an isochron with a $\sim 7 \times 10^{-5}$ slope that could then represent the $^{26}\text{Al}/^{27}\text{Al}$ ratio of their precursors formed by condensation. Other recent high precision Mg isotopic analyses have contested this view with bulk CAIs having a $^{26}\text{Al}/^{27}\text{Al}$ of $5.83 (\pm 0.11) \times 10^{-5}$ (Bizzarro et al. 2004, 2005a). This might however be an “average age” for several processes spanning a few hundred thousand years since it has been confirmed by high precision and high resolution laser ablation analyses that the oldest high temperature event recorded by CAIs correspond to a “supra-canonical” $^{26}\text{Al}/^{27}\text{Al}$ ratio of 7×10^{-5} (Young et al. 2005).

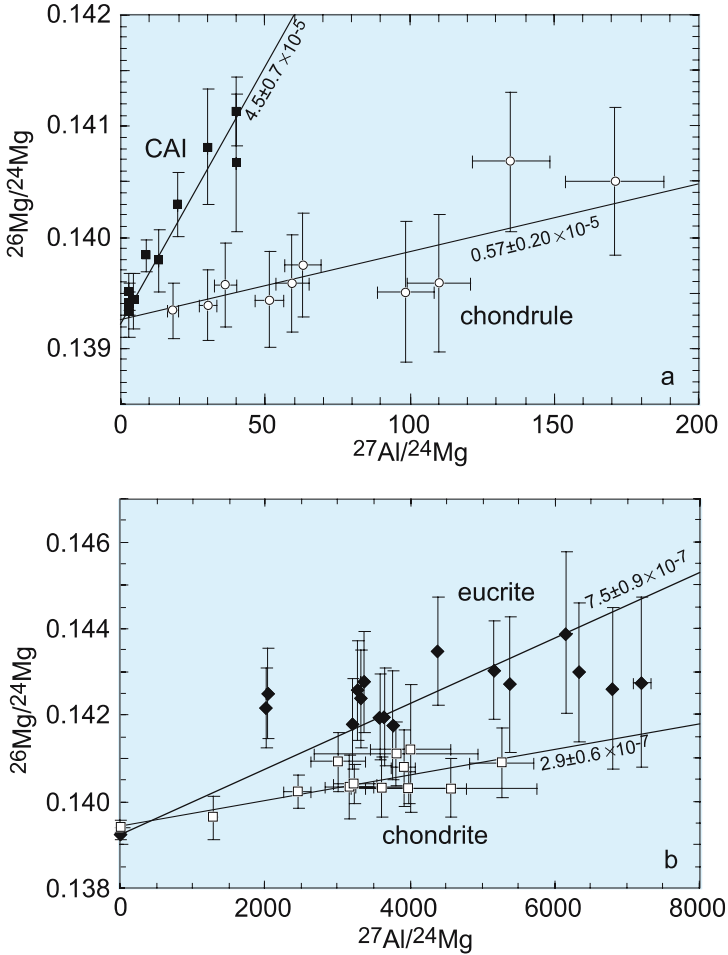


Fig. 2.2. Isochron diagrams for the $^{26}\text{Al}/^{26}\text{Mg}$ system. The *slopes of the lines* give the $^{26}\text{Al}/^{27}\text{Al}$ ratio that is time-dependent. An example of CAI (black square, Podosek et al. 1991) and chondrules (Mostéfaoui et al. 2002) is given in (a). The detections of ^{26}Al in a differentiated eucrite (Srinivasan et al. 1999) and in metamorphosed feldspaths in chondrites (Zinner and Göpel 2002) is shown in (b). The ^{26}Al indicates fast condensation and early (few Ma after CAIs) differentiation of planetesimals

Thanks to the analytical advances of ion probe microanalysis, ^{26}Mg excesses have been identified in many chondrules (Fig. 2.2a) having localized $^{27}\text{Al}/^{24}\text{Mg}$ values favorable for the detection of such excesses (e.g., Russell et al. 1996; Mostéfaoui et al. 2002). The Al/Mg fractionation observed in chondrules was produced during the partial crystallization of these silicate liquid droplets in the nebula. The initial $^{26}\text{Al}/^{27}\text{Al}$ ratios of chondrules vary between 1.5×10^{-5} and 0.3×10^{-5} (see synthesis of Wadhwa and Russell 2000). In the chronological

interpretation of ^{26}Al , this would correspond to a time difference of about 3Ma between the condensation of the CAIs and that of the precursors of the chondrules. The U/Pb absolute dating of CAIs in the chondrite CV Efremovka and of chondrules from the CR Acfer chondrite provides a solid confirmation of this interpretation (Amelin et al. 2002). Very precise measurements have yielded an age of $4567.2 \pm 0.6\text{Ma}$ for the CAIs (previously CAIs of the CV Allende chondrite gave an age of 4566^{+2}_{-1}Ma , Manhès et al. 1988; Allègre et al. 1995) and $4564.7 \pm 0.6\text{Ma}$ for the chondrules. The $2.5 \pm 1.2\text{Ma}$ age difference agrees well with the chronological interpretation of ^{26}Al . Nevertheless, this difference in age between CAI and chondrules has not yet been firmly established. Recently some chondrules in CV3 chondrites seem to have ages, within errors, close to those of the CAIs from the same meteorites (Amelin et al. 2004). Also, $^{26}\text{Al}/^{27}\text{Al}$ initial ratios of $\sim 5 \times 10^{-5}$ (identical to those of the CAIs) were found in chondrules (Galy et al. 2000; Bizzarro et al. 2004). This indicates either that the chondrules effectively underwent a melting episode at the same time as the CAIs or that their precursors were contemporaneous with those of CAIs. It is important to note here that the chronological information obtained depends strongly on the type of analysis. With the ion probe or laser ablation-ICPMS analysis, single phases (minerals or glass), which based on petrography obviously co-crystallized, can be analyzed with a resolution of $\sim 20\text{--}50\mu\text{m}$. Classical ICPMS analyses even of small samples always provide a bulk measurement. The first type of analysis may give access to a crystallization age because the fractionation measured for the Al/Mg ratios is inherited from the crystallization. The second may give access to the age of the fractionation responsible for the bulk Al/Mg ratio of the chondrule or the CAI: this bulk Al/Mg ratio can be established during equilibrium of the precursors of CAIs and chondrules with the nebular gas. In addition, it seems that younger chondrules do exist, as shown by recent U/Pb dating of chondrules from two CB chondrites (Krot et al. 2005), which have ages of $5462.7 \pm 0.9\text{Ma}$. The origin of chondrules is still a major unsolved issue in cosmochemistry (see recent review by Zanda 2004) and it is quite possible that all chondrules do not have the same nebular origin; some of them may have formed much later than CAIs during collisions between planetesimals (Krot et al. 2005).

A $^{26}\text{Al}/^{27}\text{Al}$ ratio of $7.5 \pm 0.9 \times 10^{-7}$ (Fig. 2.2b) was found in plagioclases from the eucrite Piplia Kalan (Srinivasan et al. 1999). This ratio is much lower than that of CAIs and chondrules and represents a time difference of $4.4 \pm 0.1\text{Ma}$ relative to the CAIs canonical value of 5×10^{-5} . The plagioclase crystals of the eucrite Piplia Kalan are interpreted to have formed during the differentiation of the eucrite parent body (maybe the asteroid Vesta). This would imply that the differentiation of Vesta (or of a same size planetesimal) occurred less than 5Ma after the condensation of the first solids in the forming solar system. If this was the case, the radioactive decay of ^{26}Al could have been a heat source responsible for this rapid differentiation (Urey 1955; Fish et al. 1960; Herndon and Herndon 1971). The observation of traces of ^{26}Al in the plagioclases of several H4 chondrites (Zinner and Göpel 2002) also supports the existence of

enough ^{26}Al in the planetesimals to allow their differentiation. The plagioclases (interpreted as secondary phases of metamorphic origin) in these chondrites have ^{26}Mg excesses (Fig. 2.2b) corresponding to $^{26}\text{Al}/^{27}\text{Al}$ ratios between $2.9 \pm 0.6 \times 10^{-7}$ and $1.5 \pm 0.5 \times 10^{-7}$ at the time of the metamorphism of the parent body. If these values were compared to the canonical $^{26}\text{Al}/^{27}\text{Al}$ value, metamorphism in the parent body of the H4 chondrites would have taken place 5.2 to 6.5 Ma after the CAIs formation. An early differentiation of some planetary bodies in the solar system has been recently demonstrated by the old U/Pb age of $4.5662 \pm 0.0001\text{Ga}$ (i.e. only 1 Ma younger than the CAIs U/Pb age) found for two basaltic angrites (Baker et al. 2005). If differentiation occurred that early, then ^{26}Al and ^{60}Fe are present in enough amounts to be the heat sources responsible for the differentiation (Bizzarro et al. 2005b).

2.4.3 A Relative Chronology Based on the Extinct Radioactivity of ^{53}Mn

Similarly to ^{26}Al , the extinct radioactivity of ^{53}Mn (^{53}Mn decays to ^{53}Cr with a half-life of 3.7 Ma) is widely used for relative chronology because excesses of the daughter isotope ^{53}Cr are detected in a very large number of primitive or differentiated objects. The traces of the in situ decay of ^{53}Mn were first discovered in CAIs (Birck and Allègre 1985, 1988). These first measurements have shown the existence of ^{53}Cr excesses caused by the decay of ^{53}Mn but they also evidenced that this system would be rather delicate to use because of the existence of nucleosynthetic isotopic anomalies impinging on the Cr isotopic composition (visible on the ^{54}Cr , Birck and Allègre 1988; Rotaru et al. 1992). By comparison with ^{26}Al , such a nucleosynthetic anomaly was never evidenced at the same scale for Mg. The origin of the elementary Mn/Cr fractionation in the various phases of a CAI is magmatic. As for ^{26}Al , “the ^{53}Mn age” of the CAIs may be considered as a condensation age if their magmatic history is short. An isochron for the ^{53}Mn ($^{53}\text{Cr}/^{52}\text{Cr}$ vs. $^{55}\text{Mn}/^{52}\text{Cr}$, Fig. 2.3) was obtained for different phases of CAIs in the BR1 of the Allende meteorite. This isochron gives a $^{53}\text{Mn}/^{55}\text{Mn}$ of $3.7 \pm 1.2 \times 10^{-5}$ in agreement with the analyses of other CAIs (Birck and Allègre 1985, 1988). When adding to these BR1 measurements, those made on spinels of an Efremovka CAI, and by deciding, arbitrarily that CAIs and the bulk rock chondrites are synchronous (Fig. 2.3a), a value of $2.8 \pm 0.3 \times 10^{-5}$ can be calculated for the $^{53}\text{Mn}/^{55}\text{Mn}$ ratio (Nyquist et al. 1999, 2001). It must be noted here that the Mn/Cr fractionation observed amongst bulk chondrites results most likely from condensation processes in the nebula that produced the precursors of the various chondritic components (Birck et al. 1999). There is thus some rationale to consider that the CAIs and the bulk rock chondrites define the same isochron. Nevertheless, it is obvious that the precursors of the CAIs and of all the components of chondrites did not condense at the same temperatures and thus at the same time. In a chronological interpretation, the range of $^{53}\text{Mn}/^{55}\text{Mn}$ ratios within CAIs could indicate a maximum duration for the

condensation of about 3.6Ma (duration given by the two isochrons of Fig. 2.3a). However, the isotopic compositions of chondrules indicate lower $^{53}\text{Mn}/^{55}\text{Mn}$ ratios of $7.3 \pm 3.5 \times 10^{-6}$ (Fig. 2.3b) as shown by chondrules from the Chainpur meteorite, for example (Nyquist et al. 2001). These results are difficult to explain because the lines defined from chondrule and CAIs data may not be real isochrons since each analysis corresponds to a different object: these isochrons could then be seen as mixing lines. By calculating a ^{53}Mn isochron associating chondrules with the bulk rock chondrites, a value of $9.5 \pm 3.0 \times 10^{-6}$ is found for the $^{53}\text{Mn}/^{55}\text{Mn}$ ratio during the chondrules formation (Nyquist et al. 2001). In a chronological interpretation, this indicates a time interval of 3.7 to 10.8Ma between the formation of chondrules and of CAIs.

The extinct radioactivity ^{53}Mn has also been studied in details to establish a chronology for the differentiation of planetesimals (Lugmair and Shukolyukov 1998 and references therein). Bulk rock eucrites define a ^{53}Mn isochron (Fig. 2.3c) with a $^{53}\text{Mn}/^{55}\text{Mn}$ ratio of $4.7 \pm 0.5 \times 10^{-6}$ (Lugmair and Shukolyukov 1998). This isochron, in theory, dates the individualization of the precursors of the eucrite parent body from the nebula. In a chronological interpretation, a time interval with the CAIs formation of ~ 11 Ma may be calculated. If each eucrite is now considered separately, ^{53}Mn isochrons are obtained by analyzing different mineral phases. Here, the process dated is the magmatic differentiation that produced each of the eucrites inside their parent body. These isochrons have very variable slopes ($^{53}\text{Mn}/^{55}\text{Mn}$ ratio) between $3.7 \pm 0.4 \times 10^{-6}$ and ~ 0 (or $< 1.2 \times 10^{-7}$), which would indicate a duration for the differentiation processes inside the parent body of ~ 18 Ma. Of course, this long time span must be considered with precaution because it may in part reflect the duration of metamorphism in the parent body. The U/Pb dating of phosphates in ordinary chondrites (phosphates have closing temperatures lower than those of the chromite and silicate used to obtain the ^{53}Mn isochrons of eucrites) has shown that about 60Ma might be estimated for the metamorphism on an undifferentiated parent body (Göpel et al. 1994). An approach followed to minimize the uncertainty on the age induced by metamorphism was to analyze the angrites. These meteorites are very special objects that have gone through an early differentiation, cooled very quickly and do not show important signs of secondary perturbation. Both angrites studied (bulk rock and mineral separates) define a very precise isochron (Fig. 2.3d) with a $^{53}\text{Mn}/^{55}\text{Mn}$ value of $1.25 \pm 0.07 \times 10^{-6}$ (Lugmair and Shukolyukov 1998). This ratio would indicate an age difference with the CAIs of ~ 18 Ma.

2.4.4 Absolute Calibration of the ^{26}Al and ^{53}Mn Chronologies

The relative chronologies established based on variations of the $^{26}\text{Al}/^{27}\text{Al}$ and $^{53}\text{Mn}/^{55}\text{Mn}$ ratios are not calibrated by absolute dates. The only way to achieve this chronologic calibration is to date precisely one of the components used to establish the relative chronology. Two opposing calibration points are used for

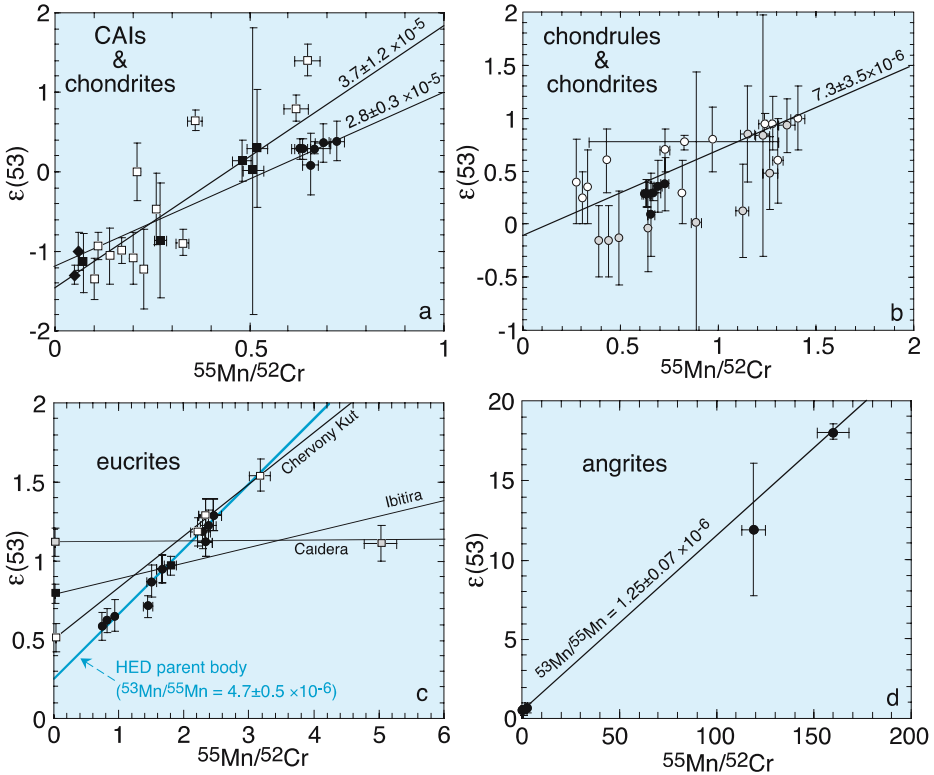


Fig. 2.3. Isochron diagram for the $^{53}\text{Mn}/^{53}\text{Cr}$ system. The variations of the Cr isotopic signature are represented as relative values compared to a reference (ϵ notation in a tenth of a thousand). The slopes of the isochrons give the $^{53}\text{Mn}/^{55}\text{Mn}$ that is time-dependent. The data for CAIs (*black and white squares*, **a**) are from Birck and Allègre (1985, 1988) and Nyquist et al. (1999) for the *black diamonds*. The *black circles* (**a** and **b**) represent bulk rock chondrites and the *white ones* chondrules (Nyquist et al. 2001). Two isochrons are visible in (a) (see text); the one obtained on the CAI BR1 (*black squares*) gives an initial maximum $^{53}\text{Mn}/^{55}\text{Mn}$ ratio for the solar system of 3.7×10^{-5} . The eucrites and angrites data (**c** and **d**) are from Lugmair and Schukolyukov (1998). Various isochrons can be drawn. They represent either the differentiation of each eucrite (*white squares* for Chervony Kut, *black squares* for Ibitira and *grey* for Caldera) or, if bulk rock data are considered, the individualization of the precursors of the eucrites parent body (*black circles*)

the ^{26}Al and the ^{53}Mn . In the case of ^{26}Al , the $^{26}\text{Al}/^{27}\text{Al}$ ratios are fixed with the maximum precision for the CAIs (canonical value of $^{26}\text{Al}/^{27}\text{Al} = 5 \times 10^{-5}$). An absolute U/Pb age of $4567.2 \pm 0.6\text{Ma}$ (Amelin et al. 2002) or of $4566 \pm 1\text{Ma}$ (Manhès et al. 1988; Allègre et al. 1995) was measured for the CAIs. In the case of ^{53}Mn , the $^{53}\text{Mn}/^{55}\text{Mn}$ ratios are anchored with the most confidence for the angrites ($^{53}\text{Mn}/^{55}\text{Mn}$ value = $1.25 \pm 0.07 \times 10^{-6}$). The angrites have the other

advantage of having been dated with great precision by the Pb/Pb method at 4557 ± 0.5 Ma (Lugmair and Galer 1992). The calibration points of both relative chronologies differ thus not only in terms of age but also in terms of processes (condensation in the nebula for ^{26}Al and differentiation in the parent body for ^{53}Mn). This anchoring of the $^{53}\text{Mn}/^{55}\text{Mn}$ chronology may be improved in the future from the recent high precision old U/Pb age (4.5662 ± 0.0001 Ga) found for two basaltic angrites (Baker et al. 2005).

2.4.5 Longer Period Extinct Radioactivities and Chronology of the Differentiation

In the last few years, the study of the extinct radioactivity of ^{182}Hf (^{182}Hf decreases into ^{182}W with a half-life of 9Ma) has seen major new developments (Lee and Halliday 1995, 1996, 1997), which deeply influenced models of planetary formation. The following chapter concentrates on this system using mostly the most recent data available for the isotopic composition of W (Quitté et al. 2000, 2005; Quitté and Birck 2004; Yin et al. 2002; Kleine et al. 2002, 2004, 2005; Schersten et al. 2006), which, for some of them, differ significantly from older data (Lee and Halliday 1995, 1996, 1997).

Because the ^{182}Hf half-life is longer than that of ^{26}Al or ^{53}Mn and because the precision of the measurements of the isotopic composition of W were recently improved, the $^{182}\text{Hf}/^{182}\text{W}$ system is very powerful to trace differentiation events lasting over a long period, i.e., from the nebular processes to the formation of the terrestrial planets. Furthermore, the $^{182}\text{Hf}/^{182}\text{W}$ allows in theory, to document both silicate-silicate and silicate-metal differentiations (mantle and crust formation as well as core and mantle formation in asteroids). Actually, Hf and W are both incompatible elements during partial melting of silicates. Since W is more incompatible than Hf, it has a much greater solubility in the liquid phase than in the crystal. During the mantle differentiation of a planetesimal, this difference of geochemical behavior induces variations of the Hf/W values in the silicate liquid (the Hf/W value is lower than that of the solid phase). In addition, W is a siderophile element (it has a strong solubility in metal) while Hf is lithophile (more soluble in the silicate liquid). Therefore, during the differentiation of a metal-rich core in a planetesimal, W and Hf will behave oppositely. The metal core will *freeze* the isotopic composition of W in the starting reservoir at the time of differentiation because the amount of ^{182}Hf incorporated in the metal is negligible compared to that of ^{182}W . The opposite effect will occur in the residual silicates with, by mass balance, a very strong increase of $^{182}\text{Hf}/^{182}\text{W}$ that will lead within a few Ma to a strong excess of ^{182}W in the silicates. The $^{182}\text{W}/^{184}\text{W}$ ratio of an object indicates the time of the metal-silicate differentiation, though a precise estimate might be model-dependent because of possible partial re-equilibration by diffusion between silicate and metal. In summary, if a silicate object has a chondritic $^{182}\text{W}/^{184}\text{W}$ signature, it implies that the metal-silicate differentiation either did not occur or that it did occur when all

the ^{182}Hf was already extinct (that is $>60\text{Ma}$ after the formation of the solar system). When the $^{182}\text{W}/^{184}\text{W}$ value of that same silicate object is superior to that of chondrites, it indicates that the metal-silicate differentiation took place when ^{182}Hf was still present.

In contrast to the $^{53}\text{Mn}/^{53}\text{Cr}$ system, Hf and W are two refractory elements (calculated temperature for 50% of condensation in equilibrium with a gas of solar composition $\approx 1690\text{K}$ for Hf and $\approx 1790\text{K}$ for W, Lodders 2003). Thus, only limited elemental fractionations are expected when the CAIs, the chondrules and/or their precursors are formed in the solar nebula (by condensation and/or volatilization processes). Two $^{182}\text{Hf}/^{182}\text{W}$ isochrons have been defined for the chondrites H4 Forest Vale and St Marguerite (Kleine et al. 2002), with slopes ($^{182}\text{Hf}/^{180}\text{Hf}$ ratio) of $1.0 \pm 0.5 \times 10^{-4}$ and of $0.85 \pm 0.05 \times 10^{-4}$, respectively (Fig. 2.4a). These isochrons were obtained by separating the magnetic minerals from non-magnetic ones of these two chondrites. Although in these chondrites, the origin of the Hf/W fractionation is not well-known, the two isochrons serve as reference points in the evolution of the solar system, in particular for the processes of differentiation. This hypothesis is strengthened by the fact that the Allende and Murchison carbonaceous chondrites (bulk rocks), as well as the CAIs in Allende (Yin et al. 2002), show in an $^{182}\text{Hf}/^{182}\text{W}$ isochron diagram, compositions rather similar from those of chondrites (Fig. 2.4a). By calculating, an isochron through all of these points, an initial $^{182}\text{Hf}/^{180}\text{Hf}$ ratio of $1.0 \pm 0.1 \times 10^{-4}$ was proposed for the solar system. A higher initial of $1.60 \pm 0.25 \times 10^{-4}$ can be inferred according to a different approach, using the lowest $^{182}\text{W}/^{184}\text{W}$ ratio measured in iron meteorites and a model chondritic evolution curve (Quitté and Birck 2004). However, recent work by Kleine et al. (2005) on CAIs has confirmed a $^{182}\text{Hf}/^{180}\text{Hf}$ of $1.07 \pm 0.10 \times 10^{-4}$ and an $\varepsilon(^{182}\text{W}/^{184}\text{W})$ of -3.47 ± 0.20 at the time of formation of CAIs, all magmatic iron meteorites having $\varepsilon(^{182}\text{W}/^{184}\text{W})$ slightly smaller or similar to that of CAIs when effects due to galactic cosmic rays spallation are taken into account (Kleine et al. 2005; Schersten et al. 2006).

The eucrites define a bulk rock $^{182}\text{Hf}/^{182}\text{W}$ isochron with a slope of $7.96 \pm 0.34 \times 10^{-5}$ (Quitté et al. 2000). If considering an initial solar system $^{182}\text{Hf}/^{180}\text{Hf}$ ratio of $1.60 \pm 0.25 \times 10^{-4}$ (this value is still discussed), the eucrite isochron would indicate a differentiation $9.5 \pm 1.5\text{Ma}$ after solar system formation. If a value of $1.0 \pm 0.1 \times 10^{-4}$ (as suggested by CAIs) is taken for the initial solar system $^{182}\text{Hf}/^{180}\text{Hf}$ ratio, eucrites would have differentiated $3 \pm 2\text{Ma}$ after the formation of the CAIs. The initial $^{182}\text{W}/^{184}\text{W}$ (0.864899 ± 49 if recalculated for a terrestrial $^{182}\text{W}/^{184}\text{W}$ ratio of 0.865000 (Quitté and Birck 2004) and $^{182}\text{Hf}/^{180}\text{Hf}$ ($7.96 \pm 0.34 \times 10^{-5}$), ratios of eucrites are above the modeled evolution of the solar system. This can be interpreted as indicating that an episode of metal-silicate differentiation has preceded that of silicate-silicate differentiation responsible for the range of variation of the $^{182}\text{Hf}/^{182}\text{W}$ signature of eucrites. In this hypothesis, the Hf/W systematic indicates that this metal-silicate differentiation (formation of the core in Vesta) occurred before the crustal differentiation

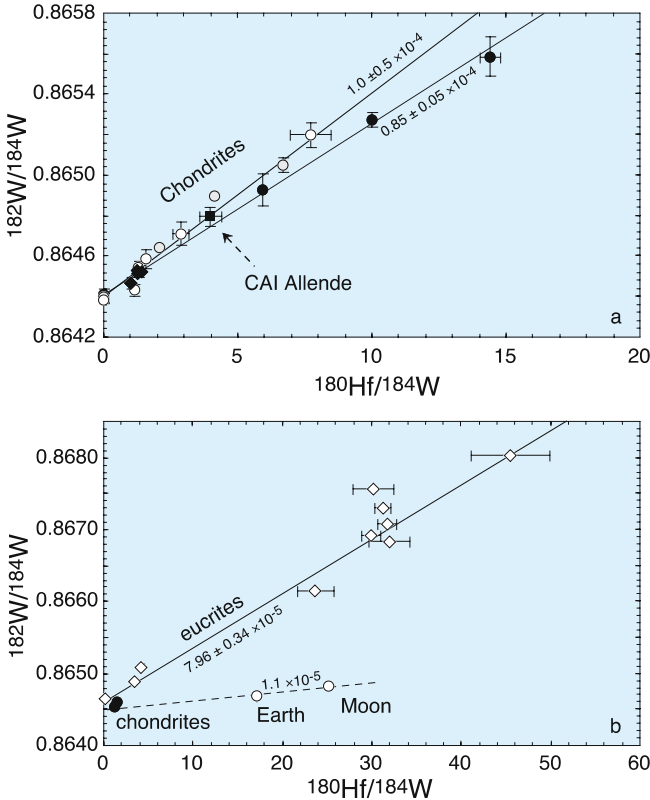


Fig. 2.4. Isochron diagrams in the $^{182}\text{W}/^{184}\text{W}$ system. The *circles* represent chondrite analyses (*black* St Marguerite, *grey* Forest Vale, *white* Dhurmsala and Dalgety), the *black square* an Allende CAI and *black diamonds* bulk rock analyses of the carbonaceous chondrites Allende and Murchison (a) (data from Kleine et al. 2002; Yin et al. 2002). Two isochrons are drawn for both St Marguerite and Forest Vale. *White diamonds* represent bulk rock eucrites (Quitté et al. 2000) compared to the bulk rock values of chondrites, the Earth and the Moon (b) (see Yin et al. 2002 for references). The eucrites define an isochron of slope $^{182}\text{Hf}/^{180}\text{Hf} = 7.96 \pm 0.34 \times 10^{-5}$ and an initial $^{182}\text{W}/^{184}\text{W}$ ratio of 0.864676 ± 29 , which, renormalized to a $^{182}\text{W}/^{184}\text{W}$ terrestrial ratio of 0.865000, represents in fact a $^{182}\text{W}/^{184}\text{W}$ ratio of 0.864899 ± 49 (see Quitté and Birck 2004 for details)

of the parent asteroid of eucrites by about 0.6 to 1.2Ma (Quitté and Birck 2004; Kleine et al. 2004).

The fact that the $^{182}\text{W}/^{184}\text{W}$ of silicates from the Earth and the Moon are more radiogenic than those of chondrites (this was not the case with the old measurements) allows to trace a tentative $^{182}\text{Hf}/^{182}\text{W}$ isochron between the chondrites representing non-differentiated material and the Earth/Moon system (Fig. 2.4b). The $^{182}\text{Hf}/^{180}\text{Hf}$ slope of this isochron is 1.1×10^{-5} (Yin et al.

2002). It indicates a period of ≈ 29 Ma for the formation of the Earth and Moon relative to that of the CAIs or a formation some 65 Ma after the origin of the solar system if an initial $^{182}\text{Hf}/^{180}\text{Hf}$ value of 1.60×10^{-4} is taken. The existing data are too limited to choose between these different interpretations. Nevertheless, it is quite clear that the higher $^{182}\text{W}/^{184}\text{W}$ ratio, around two tens of thousands (2ε units), of the Earth silicates compared to chondrites implies that the terrestrial core differentiated very early in the history of the solar system, before all the ^{182}Hf had decayed. In a simple model, the Hf/W data suggests that on Earth, the metal core differentiated in about 30 Ma. This age can also be interpreted either as an average age of the metal in the planetary embryos, which formed the Earth precursors or as an average age of the differentiation processes. Based on the existing models of isotopic exchange and metal segregation during the metal-silicate differentiation, it is impossible to totally exclude, using the Hf/W systematic, that a fraction of the terrestrial core did not differentiate much later (maybe 100 Ma after formation of the solar system), however, this hypothesis seems rather unlikely.

2.5 Remaining Questions

2.5.1 Disparities Between the Various Chronologies ^{26}Al , ^{53}Mn and ^{182}Hf

Even if at first glance, all the different extinct radioactivities (e.g., ^{26}Al , ^{53}Mn or ^{182}Hf) provide similar chronologies for condensation and differentiation processes in the early solar system, in details, major inconsistencies, superior to the error bars inherent to each isotopic system, are present. Figure 2.5 compares the two chronologies derived from ^{26}Al and ^{53}Mn .

The first and most evident inconsistency concerns the age of the CAIs and by extension that of the formation of the solar system. According to the ^{53}Mn chronology, calibrated by the Pb/Pb age of angrites, the CAIs would be 4575.4 ± 1.6 Ma old (Lugmair and Shukolyukov 1998). This age is about 8 Ma older than that obtained by U/Pb for the CAIs (Manhès et al. 1988; Amelin et al. 2002). This “too young” U/Pb age measured for CAIs cannot be explained by metamorphic perturbation of the U/Pb system, because the CAIs cooled very quickly in the nebula. One explanation may be related to the existence of Cr isotopic anomalies (visible on ^{54}Cr , e.g., Rotaru et al. 1992), which suggest that the hypothesis of an isotopic homogeneity of Cr ($^{53}\text{Cr}/^{52}\text{Cr}$) may not be correct. The $^{53}\text{Mn}/^{55}\text{Mn}$ isotopic homogeneity of the solar system in formation may also be questioned, because a fraction or all the ^{53}Mn might have been produced by irradiation processes inside the early solar system (see Sect. 2.5.2). Alternatively, as mentioned above (see Sect. 2.4.4), the anchoring of the $^{53}\text{Mn}/^{55}\text{Mn}$ has to be re-evaluated in light of the recent U/Pb ages for basaltic angrites.

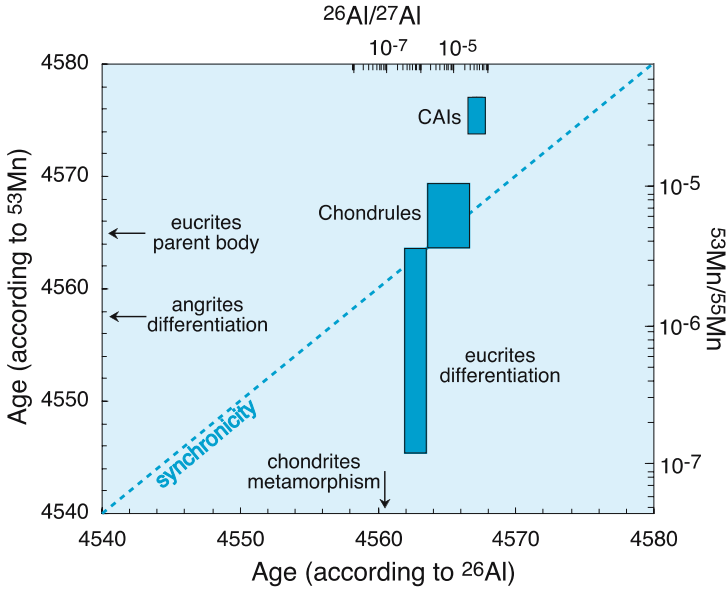


Fig. 2.5. Comparison between relative chronologies indicated by the $^{26}\text{Al}/^{26}\text{Mg}$ and $^{53}\text{Mn}/^{53}\text{Cr}$ systems. As written in the text, even if the relative order is respected for the different objects in both chronologies, important inconsistencies are obvious, in particular for the age of the CAI or the duration of the differentiation of the eucrites. These issues are under scrutiny by many groups and might be rapidly cleared from the combination of constraints brought by long-lived and short-lived radioactivities

The second discrepancy concerns the age and the duration of the differentiation of eucrites. The ^{53}Mn chronology indicates an extended differentiation period for the eucrites (lasting up to 18Ma). The detection of ^{26}Al in the Pilia Kalan eucrite advocates only a maximum of 4.4Ma after the formation of the CAIs. In the eucrites, the ^{26}Al and ^{53}Mn chronologies agree on the absolute age proposed for the beginning of the differentiation but, once again, this may be pure coincidence because the relative ages compared to the CAIs do not match. The differentiation would start 4.4Ma (based on ^{26}Al) and 12Ma (based on ^{53}Mn) after the CAIs. The ^{182}Hf systematic indicates either a differentiation of the eucrites $3 \pm 2\text{Ma}$ after the CAIs, in very good agreement with ^{26}Al or a differentiation $9.5 \pm 1.5\text{Ma}$ after the solar system formation (using the initial ratio $^{182}\text{Hf}/^{180}\text{Hf}$ deduced from iron meteorites), which would then be compatible with the ^{53}Mn estimation. If the eucrites differentiated about 10Ma after the CAIs, then neither ^{26}Al nor ^{60}Fe (half-life 1.5Ma, Shukolyukov and Lugmair 1993; Tachibana and Huss 2003; Mostéfaoui et al. 2004) can be the heat source of this differentiation. In addition, the latest high precision bulk Mg isotopic data (Bizzarro et al. 2005b) show that differentiation of the eucrite parent body and of the mesosiderite parent body occurred within 3 Ma of CAI formation. In gen-

eral, ^{26}Al seems to indicate shorter durations than the other chronometers. For example, it implies that the metamorphism on the parent body of the ordinary chondrites, St Marguerite and Forest Vale, ended between 5.2 and 6.5Ma after the formation of CAIs. This discrepancy exists also when the ages of the CAIs and the chondrules are compared. Between the chondrules and the CAIs, ^{26}Al indicates a time difference in the order of $2.1 \pm 1.5\text{Ma}$ that seems consistent with the rather short duration of the nebular processes (based on astrophysical considerations) while the ^{53}Mn suggests a 9Ma duration.

2.5.2 The Hypothesis of Homogeneity of the Distribution of Extinct Radioactivities in the Solar Accretion Disk: the Origin of Extinct Radioactivities

Any chronological interpretation of the systematic of the isotopic variations of an extinct radioactivity relies on major hypotheses regarding its initial distribution in the accretion disc. The classical (or canonical) hypothesis assumes the isotopic homogeneity of the nebula or the accretion disk at reference time 0. However, a relative chronology can also be defined in the case on a heterogeneous distribution of ^{26}Al and ^{53}Mn (Gounelle and Russell 2005). Considering the obvious inconsistencies between the various chronologies (^{26}Al and ^{53}Mn for instance), the hypothesis of an isotopic homogeneity must be evaluated in light of the available data. In the case of ^{26}Al , the canonical $^{26}\text{Al}/^{27}\text{Al}$ ratio of 5×10^{-5} found in the CAIs of ordinary chondrites is undistinguishable from the one observed previously in the CAIs of carbonaceous chondrites. This similarity can be considered as a strong argument in favor of a homogeneous distribution of ^{26}Al in the nebula, since ordinary and carbonaceous chondrites originate from parent bodies formed at different heliocentric distances (Russell et al. 1996). Nevertheless, it has also been shown that all the CAIs had similar ^{16}O enrichments. These enrichments could suggest their formation from a common source; a peculiar source since these enrichments are not present (excepted in some chondrules) in the chondrites and the differentiated meteorites (McKeegan et al. 1998). How does one explain this observation? The CAIs might have been produced locally from a distinct reservoir in the protosolar nebula. If this was the case, arguments are lacking to establish firmly the homogeneity of the ^{26}Al distribution. This problem has not yet been solved but the internal consistency of the Al-Mg systematics derived from recent high precision analyses of CAIs, of primitive and differentiated meteorites and of the Earth (Bizzarro et al. 2005b) may be viewed as a strong argument in favour of an homogeneous distribution of ^{26}Al .

This question of the homogeneity of the distribution of extinct radioactivities in the protosolar nebula is related to the nucleosynthetic origin of the radioactive elements. The presence of ^{26}Al in CAIs very clearly shows that this isotope was present in the protosolar nebula and, thus, that nucleosynthetic events occurred slightly before or contemporaneously with the formation of CAIs, i.e., within a period preceding at maximum only by a few half-lives of ^{26}Al (0.75Ma) the

formation of CAIs. The canonical hypothesis is that this event was the explosion of a supernova (e.g., Cameron 1995 and references therein; Goswami and Vanhala 2000). In recent models which consider that the Sun birth took place in a crowded stellar environment such as HII regions observed in Orion (Hester and Desh 2005) the solar system is at proximity of one or several supernova. According to this hypothesis, the products of the supernova are injected in the solar accretion disk; the newly produced ^{26}Al (and ^{27}Al) is mixed and homogenized with the ^{27}Al (stable) already present in the cloud to yield the $^{26}\text{Al}/^{27}\text{Al}$ ratio of 5×10^{-5} . The presence of ^{60}Fe in the most primitive meteorites provides an unquestionable proof for the injection of stellar products in the forming solar system (Birck and Lugmair 1988; Tachibana and Huss 2003; Mostéfaoui et al. 2004). In fact, ^{60}Fe is a neutron-rich nuclide that could not have been produced by the spallation reactions considered to have occurred when the solar system formed (Lee et al. 1998). On the contrary, ^{60}Fe is very efficiently produced by supernova and significantly, but with less efficiency, by the AGB stars (Busso et al. 2003). However, it is not possible to precisely model from the abundance of ^{60}Fe in meteorites the time ($\approx 1\text{Ma}$ or slightly less) that separates the supernova explosion (or the end of the life of the AGB star) and the isolation of the pre-solar nebula from the interstellar environment. This depends on the production rates of ^{26}Al or of ^{60}Fe in the supernova and on the amount of mixing between the products of the supernova and the pre-solar cloud (in most models on the order of 10^{-4} , e.g., in Busso et al. 2003). It is important to stress that the existing models do not predict a perfect homogeneity of the products of the supernova within the pre-solar cloud (Vanhala and Cameron 1998). In addition, the absence of ^{26}Al in hibonite grains from the Murchison meteorite, (very primitive refractory minerals having nucleosynthetic anomalies in neutron-rich isotopes such as ^{48}Ca and the ^{50}Ti), is taken as evidence that these grains condensed in internal zones of the pre-solar cloud, before the arrival of extinct radioactivities injected from the supernova (Sahijpal et al. 2000). Finally, it is worth noting that the most recent Ni isotopic data on meteorites suggest that ^{60}Fe may have been injected in the solar system $\approx 600\,000$ after the formation of CAIs (Bizzarro et al. 2006).

The discovery of the extinct radioactivity of ^{10}Be (^{10}Be decays to ^{10}B with a half-life of 1.5Ma) in CAIs (McKeegan et al. 2000; Sugiura et al. 2001; MacPherson et al. 2003) has revived the problem of the origin of the extinct radioactivities. Spallation reactions (collisions between protons or accelerated alpha particles and O and C nuclei), and not stellar processes, produce the ^{10}Be . Its presence at the time of the formation of the CAIs supports the existence of irradiation processes around the young Sun in formation and the incorporation in the CAIs or in their precursors of the products of these reactions. Alternatively, ^{10}Be might result from the magnetic trapping of galactic cosmic rays (GCRs) in the molecular cloud parental to the solar system (Desch et al. 2004). However, isotopic variations of Li and B (the only other elements produced in significant amounts by Be irradiation processes) in chondrules show that their

precursors also contain a significant fraction of irradiation products (Chaussidon and Robert 1995, 1998). An excess of ${}^7\text{Li}$ in the Allende CAI suggests a tight temporal link between irradiation around the young Sun and the formation of the CAIs (Chaussidon et al. 2006). This excess is very probably due to the in situ decrease of short-lived ${}^7\text{Be}$ (${}^7\text{Be}$ decays to ${}^7\text{Li}$ with a half-life of 53 days). In fact, the very short half-life of ${}^7\text{Be}$ dictates that the CAI was formed and irradiated at the same time and same place than ${}^7\text{Be}$ was made. Recent astrophysical observations (Feigelson and Montmerle 1999; Montmerle 2001, 2002) show that young stars, in their pre-T Tauri phase go through a stage of intense X activity (on the order of 10^5 times more intense than the actual Sun) which can last up to $\approx 10\text{Ma}$ (Feigelson et al. 2002) and during which an important irradiation of part of the accretion disk is possible. Models (Shu et al. 1994, 1997) developed to account for the magnetic interaction between the accretion disk and the star predict the existence of a reconnection zone of the magnetic field near the star (at about 0.06 AU). The X wind model developed by Frank Shu and his coworkers considers that some of the particles and the gas accreting onto the Sun might be driven into the reconnection zone, where they can be irradiated close to the Sun, and then periodically extracted from this zone by fluctuations of the magnetic field and launched by the X wind to asteroidal distances. In such a model, the precursor grains are irradiated before being assembled and melted to form the CAIs and chondrules (Shu et al. 1996, 2001). The irradiation calculations according to the X wind model show that the extinct radioactivities detected in the CAIs (${}^7\text{Be}$, ${}^{10}\text{Be}$, ${}^{26}\text{Al}$, ${}^{41}\text{Ca}$ and ${}^{53}\text{Mn}$) might be produced in abundances close to the ones measured in the CAIs (Lee et al. 1998; Gounelle et al. 2001, 2006). Of course, these models have their own uncertainties (Chaussidon and Gounelle 2006), but they demonstrate that an internal (within the solar system) origin by irradiation of the extinct radioactivities is a possible scenario. In such a model, the abundances of the various extinct radioactivities depend on several parameters such as the flux of accelerated particles and the duration of the irradiation, the cross-sections of the different reactions considered but also on the chemical composition of the target (see Chaussidon and Gounelle 2006). It is then possible (or even likely) that the different components of the meteorites have incorporated different fractions of some of the extinct radioactivities as they were formed by irradiation. In such a case, the hypothesis of a time 0 with a homogeneous isotopic composition in the protosolar nebula would not be fulfilled. This scenario could explain some of the chronological inconsistencies observed between the different extinct radioactivities.

2.6 Conclusions

It would obviously be presumptuous to claim that the variations of the concentrations of extinct radioactivities in meteorites provide a definitive chronology

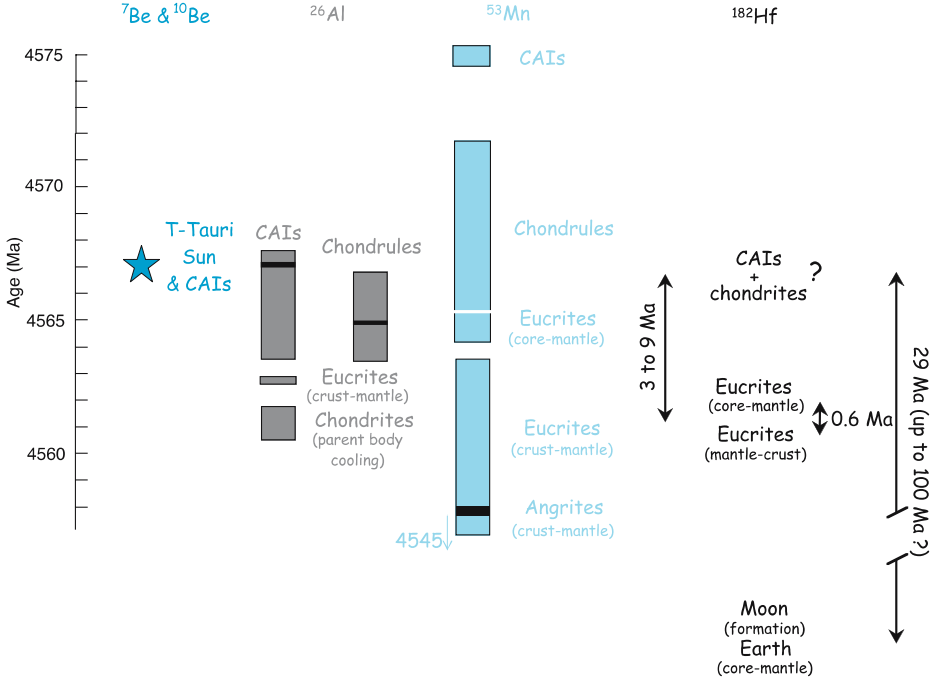


Fig. 2.6. Summary of constraints from ${}^{26}\text{Al}$, ${}^{53}\text{Mn}$, ${}^{182}\text{Hf}$ and Be isotopes on the chronology for the evolution of the early solar system. The irradiation-produced short-lived Be isotopes ties the T-Tauri phase of the young Sun and the formation of CAIs. Though discrepancies are obvious between different systems, there are strong indications that some planetary objects were accreted and differentiated very early, within ≈ 1 Ma of CAIs formation

for the formation and the early evolution of the solar system. In fact, the various extinct radioactivities show rather contrasted chronologies (Fig. 2.6). The order of magnitude of the time required to go from the gas to the grains and to the differentiated planetesimals is about ≈ 10 Ma. Shorter durations (≈ 5 Ma) are indicated by the systematic of ${}^{26}\text{Al}$. Though each isotopic chronometer possesses its own limitations, they all concur to challenge the “classical” view that chondrites represent the oldest and most primitive rocks of the solar system. Most recent U-Pb, Al-Mg and Hf-W data show that some planetary objects were already accreted and differentiated at the time (or within ≈ 1 Ma) CAIs formed. Chondrites may have been accreted later, when ${}^{26}\text{Al}$ and ${}^{60}\text{Fe}$ heat sources were nearly exhausted, and thus remained un-differentiated. More traces of this new scenario are being found. Recently, lithic clasts composed of metal-bearing olivines with granoblastic textures were identified in some chondrules (Libourel and Krot 2006): these clasts are most probably the fragments of planetary objects which were differentiated before than chondrites accreted.

Thanks

I would like to thank Jeanne Blichert-Toft, Matthieu Gounelle, Hervé Martin, Muriel Gargaud and Philippe Claeys for our discussions on solar system chronology and a careful re-reading of this manuscript.

References

- Albarède F. (2003) *Geochemistry: an Introduction*. Cambridge University Press, Cambridge
- Albarède F., Sherer E.E., Blichert-Toft J., Rosing M., Simionovici A., Bizzarro M. (2006) g-ray irradiation in the early Solar System and the conundrum of the ^{176}Lu decay constant. *Geochim. Cosmochim. Acta* **70**, 1261–1270
- Allègre C.J. (2005) *La Géologie Isotopique*. Editions Belin, Paris
- Allègre C.J., Manhès G., Göpel C. (1995) The age of the Earth. *Geochim. Cosmochim. Acta* **59**, 1445–1456
- Amelin Y., Davis W.J. (2005) Geochemical test for branching decay of ^{176}Lu . *Geochim. Cosmochim. Acta* **69**, 465–473
- Amelin Y., Krot A.N., Hutcheon I.D., Ulyanov A.A. (2002) Lead isotopic ages of chondrules and calcium-aluminum-rich inclusions. *Science* **297**, 1678–1683
- Amelin Y., Krot A.N., Twelker E. (2004) Duration of chondrule formation interval: a Pb isotopic study. In: *Proceedings of Planetary timescales: from Stardust to Continents*, Australian Academy of Science, Canberra, Australia, 16–19 February 2004
- Baker J., Bizzarro M., Wittig N., Connelly J., Haack H. (2005) Early planetesimal melting from an age of 4.5562 Gyr for differentiated meteorites. *Nature* **436**, 1127–1131
- Begemann F., Ludwig K.R., Lugmair G.W., Min K.W., Nyquist L.E., Patchett P.J., Renne P.R., Shih C.Y., Villa I.M., Walker R.J. (2001) Call for an improved set of decay constants for geochemical use. *Geochim. Cosmochim. Acta* **65**, 111–121
- Birck J.L. (2004) An overview of isotopic anomalies in extraterrestrial materials and their nucleosynthetic heritage. In: Johnson C.M., Beard B.L., Albarède F. (eds.) *Geochemistry of Non-traditional Stable Isotopes*, vol. 56, pp. 25–64. Mineralogical Society of America, Chantilly, VA
- Birck J.-L., Allègre C.J. (1985) Evidence for the presence of Mn-53 in the early solar system. *Geophys. Res. Lett.* **12**, 745–748
- Birck J.-L., Allègre C.J. (1988) Manganese-chromium isotope systematics and the development of the early solar system. *Nature* **331**, 579–584
- Birck J.-L., Lugmair G.W. (1988) Nickel and chromium isotopes in Allende inclusions. *Earth Planet. Sci. Lett.* **90**, 131–143
- Birck J.-L., Rotaru M., Allegre C.J. (1999) ^{53}Mn - ^{53}Cr evolution of the early solar system. *Geochim. Cosmochim. Acta* **63**, 4111–4117
- Bizzarro M., Baker J.A., Haack H. (2004) Mg isotope evidence for contemporaneous formation of chondrules and refractory inclusions. *Nature* **431**, 275–278

- Bizzarro M., Baker J.A., Haack H. (2005a) Mg isotope evidence for contemporaneous formation of chondrules and refractory inclusions – Corrigendum. *Nature* **435**, p 1280
- Bizzarro M., Baker J.A., Haack H., Lundgaard K.L. (2005b) Rapid timescales for accretion and melting of differentiated planetesimals inferred from ^{26}Al - ^{26}Mg chronometry. *Ap. J.* **632**, L41–L44
- Bizzarro M., Ulfbeck D., Trinquier A., Thrane K., Connelly J.N. (2006) Nickel isotope anomalies in meteorites. *Meteor. Planet. Sci.* **41**, 69th Meeting of the Meteoritical Society, abstract #5217
- Blichert-Toft J., Boyet M., Télouk P., Albarède F. (2002) ^{147}Sm - ^{143}Nd and ^{176}Lu - ^{176}Hf in eucrites and the differentiation of the HED parent body. *Earth Planet. Sci. Lett.* **204**, 167–181
- Busso M., Gallino R., Wasserburg G.J. (2003) Short-lived nuclei in the early solar system: a low mass stellar source. *Astron. Soc. Australia* **20**, 356–370
- Cameron A.G.W. (1995) The first ten million years in the solar nebula. *Meteoritics* **30**, 133–161
- Chaussidon M., Gounelle M. (2006) Irradiation processes in the early solar system. In: D. Lauretta, Leshin L. (eds.) *Meteorites and Early Solar System II*. Arizona University Press, Tuscon, AZ, pp 323–339
- Chaussidon M., Robert F. (1995) Nucleosynthesis of ^{11}B -rich boron in the pre-solar cloud recorded in meteoritic chondrules. *Nature* **374**, 337–339
- Chaussidon M., Robert F. (1998) $^7\text{Li}/^6\text{Li}$ and $^{11}\text{B}/^{10}\text{B}$ variations in chondrules from the Semarkona unequilibrated chondrite. *Earth Planet. Sci. Lett.* **164**, 577–589
- Chaussidon M., Robert F., McKeegan K.D. (2006) Li and B isotopic variations in an Allende CAI: evidence for the in situ decay of short-lived ^{10}Be and for the possible presence of the short-lived nuclide ^7Be in the early solar system. *Geochim. Cosmochim. Acta.* **70**, 224–245
- Desch S.J., Connolly H.C., Srinivasan G. (2004) An interstellar origin for the beryllium-10 in calcium-rich, aluminum-rich inclusions. *Astrophys. J.* **602**, 528–542
- Faure G. (1986) *Principles of Isotope Geology*. Wiley, New York
- Feigelson E.D., Montmerle T. (1999) High energy processes in young stellar objects. *Ann. Rev. Astron. Astrophys.* **37**, 363–408
- Feigelson E.D., Garmire G.P., Pravdo S.H. (2002) Magnetic flaring in the pre-main sequence sun and implications for the early solar system. *Astrophys. J.* **572**, 335–349
- Fish R.A., Goles G.G., Anders E. (1960) The record in the meteorites: III. On the development of meteorites in asteroidal bodies. *Astrophys. J.* **132**, 243–258
- Galy A., Young E.D., Ash R.D., O’Nions R.K. (2000) The formation of chondrules at high gas pressures in the solar nebula. *Science* **290**, 1751–1753
- Galy A., Hutcheon I.D., Grossman L. (2004) $(^{26}\text{Al}/^{27}\text{Al})_0$ of the solar nebula inferred from Al-Mg systematic in bulk CAIs from CV3 chondrites. *Lun. Planet. Sci. Conf.* **XXXV**, 1790
- Göpel C., Manhès G., Allègre C.J. (1994) U-Pb systematics of phosphates from equilibrated ordinary chondrites. *Earth Planet. Sci. Lett.* **121**, 153–171
- Goswami J.N., Vanhala H.A.T. (2000) Extinct radionuclides and the origin of the solar system. In: Manning M., Boss A., Russell S. (eds.) *Protostars and Planets IV*, pp. 963–994. Arizona University Press, Tuscon, AZ

- Gounelle M., Russell S.S. (2005) On early solar system chronology: implications of an heterogeneous spatial distribution of ^{26}Al and ^{53}Mn . *Geochim. Cosmochim. Acta* **69**, 3129–3144
- Gounelle M., Shu F.H., Shang H., Glassgold A.E., Rehm E.K., Lee T. (2001) Extinct radioactivities and protosolar cosmic-rays: self-shielding and light elements. *Astrophys. J.* **548**, 1051–1070
- Gounelle M., Shu F.H., Shang H., Glassgold A.E., Rehm K.E. and Lee T. (2006) The irradiation origin of beryllium radioisotopes and other short-lived radionuclides. *Ap. J.* **640**, 1163–1170
- Hamouda T., Pichavant M., Chaussidon M. (1996) Isotopic equilibration during partial melting: an experimental test of the behavior of Sr. *Earth Planet. Sci. Lett.* **144**, 109–121
- Herndon J.M., Herndon M.A. (1977) Aluminum 26 as a planetoid heat source in the early solar system. *Meteoritics* **12**, 459–465
- Hester J.J. and Desch S. (2005) Understanding our origins; star formation in HII region environments. In: *Chondrites and the protoplanetary disk* (A.N. Krot, E.R. D. Scott and B. Reipurth (eds.)) *ASP Conf series*, vol 341, pp 107–130
- Hewins R.H. (1997) Chondrules. *Ann. Rev. Earth Planet. Sci.* **25**, 61–83
- Jeffery P.M., Reynolds J.H. (1961) Origin of excess Xe_{129} in stone meteorites. *J. Geophys. Res.* **66**, 3582–3583
- Kelly W.R., Wasserburg G.J. (1978) Evidence for the existence of ^{107}Pd in the early solar system. *Geophys. Res. Lett.* **5**, 1079–1082
- Kleine T., Münker C., Metzger K., Plame H. (2002) Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf-W chronometry. *Nature* **418**, 952–955
- Kleine T., Mezger K., Münker C., Palme H., Bischoff A. (2004) ^{182}Hf - ^{182}W isotope systematics of chondrites, eucrites, and Martian meteorites: chronology of core formation and early mantle differentiation in Vesta and Mars. *Geochim. Cosmochim. Acta* **68**, 2935–2946
- Kleine T., Mezger K., Palme H., Scherer E., Münker C. (2005) Early core formation in asteroids and late accretion of chondrite parent bodies: evidence from ^{182}Hf - ^{182}W in CAIs, metal-rich chondrites and iron meteorites. *Geochim. Cosmochim. Acta* **69**, 5805–5818
- Krot A.N., Amelin Y., Cassen P., Meibom A. (2005) Young chondrules in CB chondrites from a giant impact in the early solar system. *Nature* **436**, 989–992
- Lee D.-C., Halliday A.N. (1995) Hafnium-tungsten chronometry and the timing of terrestrial core formation. *Nature* **378**, 771–774
- Lee D.-C., Halliday A.N. (1996) Hf-W isotopic evidence for rapid accretion and differentiation in the early solar system. *Science* **274**, 1876–879
- Lee D.-C., Halliday A.N. (1997) Core formation on Mars and differentiated asteroids. *Nature* **388**, 854–857
- Lee T., Papanastassiou D.A., Wasserburg G.J. (1976) Demonstration of ^{26}Mg excess in Allende and evidence for ^{26}Al . *Geophys. Res. Lett.* **3**, 109–112
- Lee T., Shu F.H., Shang H., Glassgold A.E., Rehm K.E. (1998) Protostellar cosmic rays and extinct radioactivities in meteorites. *Astrophys. J.* **506**, 898–912
- Libourel G., Krot A.N. (2006) Origin of olivines in Type I chondrules: petrologic and chemical constraints. *Lunar Planet. Sci.* **XXXVII**, abstract #1334

- Lodders K. (2003) Solar system abundances and condensation temperatures of the elements. *Astrophys. J.* **591**, 1220–1247
- Lugmair G., Galer S.J.G. (1992) Age and isotopic relationships among the angrites Lewis Cliff 86010 and Angra Dos Reis. *Geochim. Cosmochim. Acta* **56**, 1673–1694
- Lugmair G., Schukolyukov A. (1998) Early solar system timescales according to ^{53}Mn - ^{53}Cr systematics. *Geochim. Cosmochim. Acta* **62**, 2863–2886
- MacPherson G.J., Huss G.R., Davis A.M. (2003) Extinct ^{10}Be in type A calcium-aluminum-rich inclusions from CV chondrites. *Geochim. Cosmochim. Acta* **67**, 3165–3179
- MacPherson G.J., Davis A.M., Zinner E.K. (1995) The distribution of aluminum-26 in the early solar system: a reappraisal. *Meteoritics* **30**, 365–386
- McKeegan K.D., Chaussidon M., Robert F. (2000) Incorporation of short-lived ^{10}Be in a calcium-aluminum-rich inclusion from the Allende meteorite. *Science* **289**, 1334–1337
- McKeegan K.D., Leshin L.A., Russell S.S., MacPherson G.J. (1998) Oxygen isotopic abundances in calcium-aluminum-rich inclusions from ordinary chondrites: implications for nebular heterogeneity. *Science* **280**, 414–418
- McKeegan K.D., Davis A.M. (2003) Early solar system chronology. In: Holland H., Turekian K. (eds.) *Treatise on Geochemistry*, vol. 1, pp. 431–460. Elsevier-Perгамon, Oxford
- Manhès G., Göpel C., Allègre C.J. (1988) Systématique U-Pb dans les inclusions réfractaires d'Allende: le plus vieux matériau solaire. *Comptes Rendus de l'ATP Planétologie*, 323–327
- Montmerle T. (2001) La formation des étoiles de type solaire: des conditions initiales pour l'origine de la vie? In: Gargaud M., Despois D., Parisot J.-P. (eds.) *L'environnement de la Terre Primitive*, pp. 31–52. Presses Universitaires de Bordeaux, Bordeaux
- Montmerle T. (2002) Irradiation phenomena in young solar-type stars and the early solar system: X-ray observations and γ -ray constraints. *New Astron. Rev.* **46**, 573–583
- Mostefaoui S., Kita N.T., Togashi S., Tachibana S., Nagahara H., Morishita Y. (2002) The relative formation age of ferromagnesian chondrules inferred from their initial aluminum 26/aluminum 27 ratios. *Meteor. Planet. Sci.* **37**, 421–438
- Mostefaoui S., Lugmair G.W., Hoppe P., El Goresy A. (2004) Evidence for live ^{60}Fe in meteorites. *New Astron. Rev.* **48**, 155–159
- Nyquist L.E., Shih C.-Y., Wiesman H., Reese Y., Ulyanov A.A., Takeda H. (1999) Towards a Mn-Cr timescale for the early solar system. *Lunar Planet. Sci.* **30**, 1604
- Nyquist L., Lindstrom D., Mittlefehldt D., C.-Y. Shih, Wiesmann H., Wentworth S., Martinez R. (2001) Manganese-chromium formation intervals for chondrules from the Bishunpur and Chainpur meteorites. *Meteor. Planet. Sci.* **36**, 911–938
- Patchett P.J., Vervoort J.D., Söderlund V., Salters V.J.M. (2004) Lu-Hf and Sm-Nd isotope systematics in chondrites and their constraints on the Lu-Hf properties of the Earth. *Earth Planet. Sci. Lett.* **222**, 29–41
- Podosek F.A., Zinner E.K., MacPherson G.J., Lundberg L.L., Brannon J.C., Fahey A.J. (1991) Correlated study of initial $^{87}\text{Sr}/^{86}\text{Sr}$ and Al-Mg systematics and petrologic properties in a suite of refractory inclusions from the Allende meteorite. *Geochim. Cosmochim. Acta* **55**, 1083–1110

- Podosek F.A., Swindle T.D. (1988a) Extinct radionuclides. In: Kerridge J., Matthews M. (eds.) *Meteorites and the Early Solar System I*, pp. 1093–1113. University of Arizona Press, Tuscon, AZ
- Podosek F.A., Swindle T.D. (1988b) Nucleocosmochronology. In: Kerridge J., Matthews M. (eds.) *Meteorites and the Early Solar System I*, pp. 114–1126. University of Arizona Press, Tuscon, AZ
- Quitté G., Birck J.-L. (2004) Tungsten isotopes in eucrites revisited and the initial $^{182}\text{Hf}/^{180}\text{Hf}$ of the solar system based on iron meteorite data. *Earth Planet. Sci. Lett.* **219**, 201–207
- Quitté G., Birck J.-L., Allègre C.J. (2000) ^{182}Hf - ^{182}W systematics in eucrites: the puzzle of iron segregation in the early solar system. *Earth Planet. Sci. Lett.* **184**, 83–94
- Richter F.M., Davis A.M., DePaolo D.J., Watson E.B. (2003) Isotope fractionation by chemical diffusion between molten basalt and rhyolite. *Geochim. Cosmochim. Acta* **67**, 3905–3923
- Rotaru M., Birck J.-L., Allègre C.J. (1992) Clues to early solar system history from chromium isotopes in carbonaceous chondrites. *Nature* **358**, 465–470
- Russell S.S., Srinivasan G., Huss G.R., Wasserburg G.J., MacPherson G.J. (1996) Evidence for widespread ^{26}Al in the solar nebula and constraints for nebula timescales. *Nature* **273**, 757–762
- Russell S.S., Gounelle M., Hutchinson R. (2001) Origin of short-lived radionuclides. *Phil. Trans. R. Soc. Lond. A* **359**, 1991–2004
- Sahijpal S., Goswami J.N., Davis A.M. (2000) K, Mg, Ti and Ca isotopic compositions and refractory trace element abundances in hibonites from CM and CV meteorites: implications for early solar system processes. *Geochim. Cosmochim. Acta* **64**, 1989–2005
- Shu F.H., Shang H., Lee T. (1996) Towards an astrophysical theory of chondrites. *Science* **271**, 1545–1552
- Shu F.H., Najita J., Ostriker E., Wilkin F., Rude S., Lizano S. (1994) Magnetocentrally driven flows from young stars and discs: I. a generalized model. *Astrophys. J.* **429**, 781–796
- Shu F.H., Shang H., Glassgold A.E., Lee T. (1997) X-rays and fluctuating X-winds from protostars. *Science* **277**, 1475–1479
- Shu F.H., Shang S., Gounelle M., Glassgold A.E., Lee T. (2001) The origin of chondrules and refractory inclusions in chondritic meteorites. *Astrophys. J.* **548**, 1029–1050
- Shukolyukov A., Lugmair G.W. (1993) Live iron-60 in the early solar system. *Science* **259**, 1138–1142
- Srinivasan G., Goswami J.N., Bhandari N. (1999) ^{26}Al in eucrite Piplia Kalan: plausible heat source and formation chronology. *Science* **284**, 1348–1350
- Srinivasan G., Ulyanov A.A., Goswami J.N. (1994) ^{41}Ca in the early solar system. *Astrophys. J. Lett.* **431**, L67–L70
- Stolper E., Paque J.M. (1986) Crystallization sequence of calcium-aluminum-rich inclusions from Allende: the effects of cooling rate and maximum temperature. *Geochim. Cosmochim. Acta* **50**, 1785–1806
- Sugiura N., Shuzou Y., Ulyanov A.A. (2001) Beryllium-boron and aluminum-magnesium chronology of calcium-aluminum-rich inclusions in CV chondrites. *Meteor. Planet. Sci.* **36**, 1397–1408
- Swindle T.D., Davis A.M., Hohenberg C.M., MacPherson G.J., Nyquist L.E. (1996) Formation time of chondrules and of Ca, Al-rich inclusions: constraints from

- short-lived nuclides. In: Hewins R., Jones R., Scott E. (eds.) *Chondrules and the Protoplanetary Disc*, pp. 77–86. Cambridge University Press, Cambridge
- Tachibana S., Huss G.R. (2003) The initial abundance of ^{60}Fe in the solar system. *Astrophys. J.* **588**, L41–L44
- Tilton G.R. (1988a) Principles of radiometric dating. In: Kerridge J., Matthews M. (eds.) *Meteorites and the Early Solar System I*, pp. 149–258. University of Arizona Press, Tucson, AZ
- Tilton G.R. (1988b) Age of the solar system. In: Kerridge J., Matthews M. (eds.) *Meteorites and the Early Solar System I*, pp. 259–275. University of Arizona Press, Tucson, AZ
- Urey H.C. (1955) The cosmic abundance of potassium, uranium and thorium and the heat balances of the Earth, the Moon and Mars. *Proc. Nat. Acad. Sci. US* **41**, 127–144
- Vanhala H.A.T., Cameron A.G.W. (1998) Numerical simulations of triggered star formation: I. Collapse of dense molecular cloud cores. *Astrophys. J.* **508**, 291–307
- Vidal P. (1998) *Géochimie*. Dunod, Paris
- Wadhwa W., Russell S. (2000) Timescales of accretion and differentiation in the early solar system. In: Manning M., Boss A., Russell S. (eds.) *Protostars and Planets IV*, pp. 995–1018. Arizona University Press, Tucson, AZ
- Wasserburg G.J. (1987) Isotopic abundances: inferences on solar system and planetary evolution. *Earth Planet. Sci. Lett.* **86**, 129–173
- Wetherill G.W. (1975) Radiometric chronology of the early solar system. *Ann. Rev. Nucl. Sci.* **25**, 283–328
- Yin Q., Jacobsen S.B., Yamashita K., Blichert-Toft J., Télouk P., Albarède F. (2002) A short timescale for terrestrial planet formation from Hf-W chronometry of meteorites. *Nature* **418**, 949–952
- Young E.D., Simon J.I., Galy A., Russell S.S., Tonui E., Lovera O. (2005) Supracanonical $^{26}\text{Al}/^{27}\text{Al}$ and the residence time of CAIs in the solar protoplanetary disc. *Science* **308**, 223–227
- Zanda B. (2004) Chondrules. *Earth Planet Sci. Lett.* **224**, 1–17
- Zinner E., Göpel C. (2002) Aluminum-26 in H4 chondrites: implications for its production and its usefulness as a fine scale chronometer for early solar system events. *Meteor. Planet. Sci.* **37**, 1001–1013

3 The Formation of Crust and Mantle of the Rocky Planets and the Mineral Environment of the Origin of Life

Francis Albarède

The existence of a solid/liquid interface is probably a favorable condition for the origin of life: being anchored to a solid surface prevents the anarchic and irreversible dispersion of organisms and the liquid state supplies nutriment and evacuates waste products. Today, the exact nature of this support remains unknown, did the first biological molecules appear on the continents or in the oceans? The composition of the primary atmosphere, as well as the planet's capability to produce reduced gas, such as hydrogen, methane and ammonia are also unknown. Erosion regularly transfers nutriment to the ocean and as such contributes to the preservation and to the evolution of life. Consequently, the existence of emerged land surfaces and their mineralogical composition must be taken into account. The plate tectonic activity that generates a variety of environments, involved in complex but qualitatively unchanging interactions is a major characteristic of planet Earth. The same seems to be the case for life. The dynamics of crust formation on the terrestrial planets, and that of the mantle, which led to its formation as well as their interactions with liquid water are an integral part of the many problems belonging to the theme of astrobiology. This chapter analyzes comparatively the composition and the dynamics of the telluric planets of the inner solar system, in particular Earth, Moon, Mars, and Vesta, four relatively well-known planets with available samples and which shelter, or could have sheltered, life forms.

3.1 Chemical and Mineralogical Structure of the Earth

The internal composition of the Earth is essentially documented by the observations of the velocity of elastic waves triggered by seismic activity (Fig. 3.1). A worldwide network of stations records the arrival time of the different sets of waves generated by seismic events. The characteristics of an incoming wave mainly reflect the encountered lithologies. Several major discontinuities of seismic velocities, at different depths have been identified: the boundary between the crust and the upper mantle (around a depth of ~ 40 km under the continents and ~ 6 km under the oceans), the two discontinuities in the transition zone between upper and lower mantle at 440 and 660 km, respectively, the limit between the lower mantle and the outer liquid core at a depth of 2900 km, and the discontinuity separating the outer and inner core.

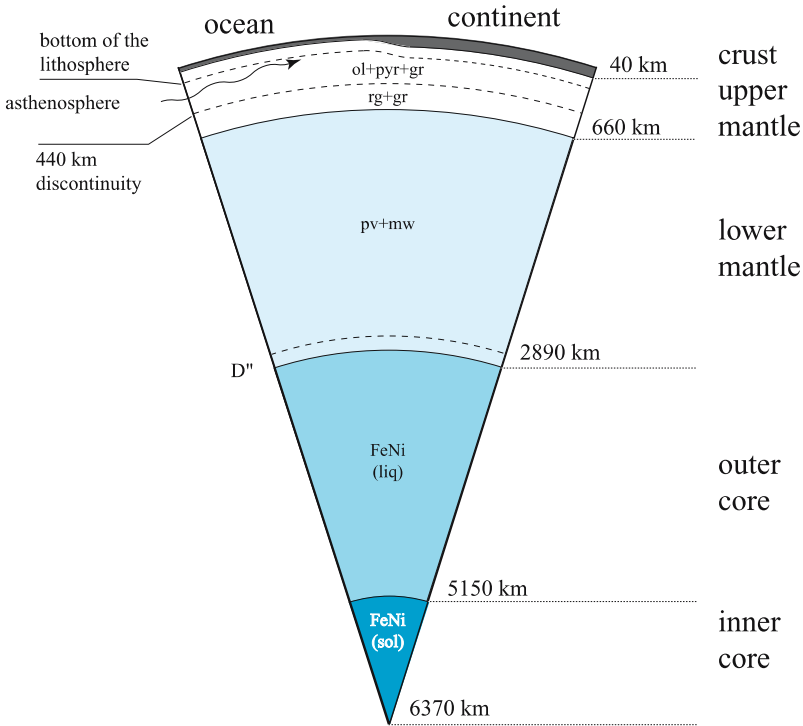


Fig. 3.1. The deep structure of the Earth as described by seismic waves. The structure matches the mineralogical composition deduced from high-pressure experiments. *Ol*: olivine, *pyr*: pyroxene, *gr*: garnet, *rg*: ringwoodite, *pv*: perovskite, *mw*: magnesio-wüstite. The nature of the D'' layer is still unclear but it may correspond to a so-called post-perovskite mineral

The propagation velocity of seismic waves reveals the physical state of the encountered material. Shear waves travel through the solid mantle but do not propagate in the liquid outer core. Seismic waves also provide information on the density of the materials, since on average, they move faster in denser rocks. Each seismic discontinuity is in fact a density jump related to a change of composition, a mineralogical reaction or, in the case of the core, a change of state of the material (solid–liquid).

The chemical and mineralogical contrast between mantle and crust is another characteristic of the Earth. The continental crust is mainly composed of rocks rich in Si, Al, Ca, K and Na, present in common minerals such as quartz, feldspars, clay minerals, and carbonates. These minerals are abundant in granitic rocks, in the sediments derived from the erosion on the crust, as well as in their metamorphic equivalents.

Under a thin layer of sediments, the oceanic crust is mainly formed of basalt, the lava resulting from the rapid crystallization of the liquid most commonly

produced by fusion of the mantle. Rocks of basaltic composition but better crystallized due to their slower cooling rate are called gabbros. The most common elements are Si, Al, Mg, Fe and Ca. These rocks are altered, especially at oceanic ridges, by the circulation of seawater within the crust, with a correlative increase in the K, Na and Si concentrations.

The composition of the upper mantle is dominated by Si, Mg and Fe and may be characterized by the contrasting properties of the minerals involved. The most refractory minerals, notably the abundant olivine $(\text{Mg,Fe})_2\text{SiO}_4$ and orthopyroxene $(\text{Mg,Fe})_2\text{Si}_2\text{O}_6$, retain most of the high melting point elements (called by geochemists, compatible elements), especially Mg and Cr. Clinopyroxene $\text{Ca}(\text{Mg,Fe})\text{Si}_2\text{O}_6$ and the aluminum-rich minerals retain the fusible elements that contribute the most to the formation of basalt (Al, Fe, Ca) particularly. The nature of these Al-rich minerals varies with depth: plagioclase $(\text{CaAl}_2\text{Si}_2\text{O}_8)$ is the stable phase between 0 and 30km, spinel $(\text{MgAl}_2\text{O}_4)$ between 30 to 60km, and then garnet $(\text{Mg}_3\text{Al}_2\text{Si}_3\text{O}_{12})$ at greater depths. The stability of these minerals plays a crucial role in the mineralogical characteristics of the different planets. The terrestrial upper mantle is thus dominated by a rock rich in olivine, known as peridotite (*peridot* being the ancient name of olivine). Its constitutive mineralogy is further referred to as garnet-, spinel- or plagioclase-peridotite, according to the nature of the aluminum-rich mineral.

With increasing depth, CO_2 reacts with silicate minerals to form carbonates. Clinopyroxene dissolves itself gradually into garnet and around 440km, olivine successively changes into its denser polymorphs, wadsleyite first, then ringwoodite. This sudden change in density constitutes the first discontinuity in the transition zone. At 660km, the second discontinuity corresponds to the destruction of ringwoodite and garnet to form perovskite and an oxide of Fe and Mg.

The core is mainly composed of the very dense metals iron and nickel, which explain the Earth's high moment of inertia. The comparison between the composition of the crust and mantle on Earth and that of the non-differentiated meteorites (chondrites) as well as the observed seismic velocities suggest that the core also contains some light elements. Up to 10% of silicon, and possibly sulfur or oxygen, are present. The movements of liquid iron, in the outer part of the core, explain the presence on Earth of a strong magnetic field.

3.2 Dynamics of the Earth's Interior

This subject was covered in detail by Sotin (2001) and only the basic principles required to understand the remaining part of the text will be discussed here. Terrestrial dynamic is regulated by the production and the dissipation of the energy available inside the Earth. First, Earth is warmed by the decay of four radioactive nuclides, isotopes of uranium, thorium, and potassium, ^{235}U , ^{238}U , ^{232}Th , and ^{40}K . U and Th alone account for 80% of the radioactive heat generated. Uranium-235, which has a short half-life ($T_{1/2}$), is actually only present

in very low concentrations. It was, however, the main source of radioactive heat shortly after the formation of the Earth. The other important energy source is the fossil heat inherited from the gravitational energy released during the accretion of the Earth. Should this energy be freed instantly, it would warm up the Earth to 35 000°C. Core formation, for its part, would only contribute an extra 1500°C to this warming. Of course, an important part of the heat radiated into space during the accretion but enough remained to account for a substantial fraction of the present-day terrestrial heat flux, which amounts to about 42TW. This flux comes almost equally from radioactive decay and fossil heat; a smaller part derives from the formation of the core and from the latent heat of its crystallization.

A warming of the deeper layers of the mantle triggers the expansion of this material against the existing pressure forces. The warmer material becomes lighter and induces a density inversion: temperature effect on density overwhelms the pressure effect and therefore deep material is warmer, thus lighter than in surface conditions. The mechanical instability generated in this way can only be compensated by mantle overturn. The deep warm and light material rises to the surface, where it cools off and decompresses. The cold dense material of the surface subsequently founders back into the deeper layers of the mantle, efficiently cooling them down in the process. This is the process of convection (Fig. 3.2). Very little mechanical energy is associated with convection because motion is very slow. The effect of convection is essentially an accelerated transfer of thermal energy from the deep to the surface.

Two modes of convection overlap in the mantle of terrestrial planets. First, internal heating by radioactive elements is homogeneously distributed throughout the mantle. This process leads to a large-scale convection with ascending and descending limbs. Heating from below, in particular upon warming of the lower mantle by the underlying core, which is several hundred degrees hotter, represents an additional mode of heating. There is no significant mass transfer across the mantle-core boundary. Therefore, most of the energy transfer occurs by conduction through a thin layer much warmer than the overlying and almost motionless mantle. Broad-scale convection is supplemented by localized instabilities (plumes or hot spots) generated in the hotter parts of this thermal boundary layer. A common opinion among the geodynamicists is that the energy conveyed by terrestrial hotspots (Hawaii, Réunion, Iceland) is equal to the heat produced in the core. This is the same energy (4–10TW), which makes possible the production of the terrestrial magnetic field by the core dynamo.

What happens to this thermal energy extracted from the depths of the Earth? Two extreme cooling mechanisms are possible. If the deep and warm material moves up to an open surface, where it cools off suddenly (by contact with water or by radiation in space), the convective regime is then controlled by surface temperature. Viscosity ν depends on temperature T according to

$$\nu = \nu_0 \exp\left(\frac{E}{RT}\right) \quad (3.1)$$

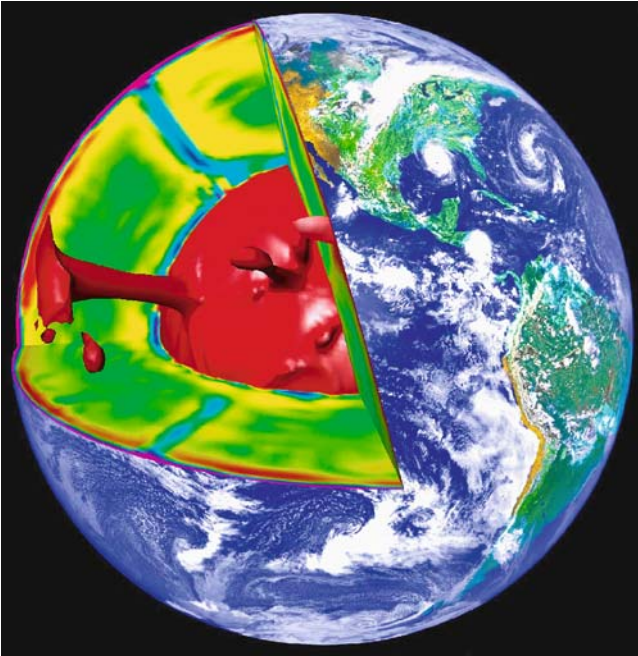


Fig. 3.2. Convection of the terrestrial mantle, as described by numerical models of convection. *Blue shading* represents the cold plates sinking into the subduction zones, and *red shading* the hotspots, which originate at the core-mantle boundary layer and are triggered by the input of heat from the core (courtesy of Paul Tackley)

where R is the universal gas constant, E the activation energy, and ν_0 a constant. When heat production increases, the mantle warms up, becomes less viscous, and therefore offers less resistance to convection. The greater the volume of deep and warm material that arrives at the surface, the higher the heat loss. In such a case, convection is regulated by thermal feedback on viscosity. The opposite extreme (Mars) is one of a planet covered by a thick and stagnant lithosphere, without plate tectonics. Heat is lost by conduction through the lithospheric lid only. Since the range of temperature variations in the mantle is limited, the heat flow at the surface remains almost constant and there is no strong relationship between heat production and energy loss at the surface. The Earth's situation falls somewhere in between these two cases (Fig. 3.3). An important part of thermal flux is extracted by the ocean along the mid-oceanic ridges where the thickness of the lithosphere is negligible. Another fraction is lost by conduction through the lithosphere, the thickness of which varies between 70km under the ocean and 250km under the continents. A moderate feedback between the heat flow and the temperature of the mantle is therefore expected. The most conspicuous demonstration of this thermal system is the existence on Earth of plate tectonics. The foundering of colder, hence denser material

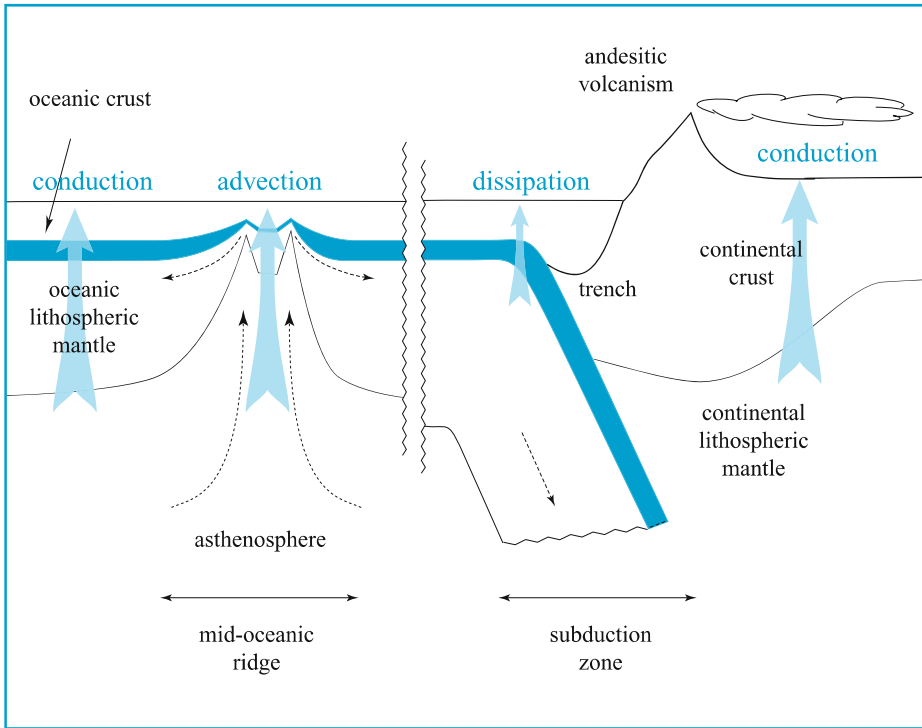


Fig. 3.3. Different paths of internal heat loss drive different types of convection in planets. When heat reaches the surface, as under mid-oceanic ridges, thermal feedback on viscosity dominates: the hotter the mantle, the lower its viscosity. As a consequence, convection becomes more vigorous and more efficient at cooling the Earth. On Mars, heat is lost by conduction through a stagnant lithospheric lid: heat loss is not much affected by mantle temperature. On Earth, the presence of water has allowed a composite situation dominated by plate tectonics. Thermal dissipation upon folding of the subducting plates is one of the important pathways for the loss of internal heat

(subduction, Fig. 3.3) is neither completely free, as in the case of a lava-lake, nor efficiently blocked, as in the case of a stagnant lithospheric lid. However, subduction is strongly slowed down by the necessity of bending the heavy, cold, and rigid tectonic plates before they are allowed to sink into subduction zones. This action dissipates an important part of the heat generated at great depths (Conrad and Hager 1999).

3.3 The Origin of Continents

This subject was discussed in detail in the previous volume, but with a slightly different perspective (Martin 2005). If the formation of oceanic crust by melting

of the underlying mantle at mid-ocean ridges seems reasonably well-understood, the formation of continents remains much of an enigma. The typical material of the continental crust, which for simplicity will be referred to as granite (although the specialist may prefer granodiorite), is formed by crystallization of silica-rich magmas. Experimental petrology demonstrates that these magmas react with the upper mantle peridotites under all depth and water-content conditions. In a dry environment, olivine transforms into orthopyroxene, while in a wet environment it forms amphibole. As a consequence, granitic rocks by themselves cannot form by straight melting of ordinary mantle rocks. The most commonly held opinion is that oceanic crust, hydrated and altered by seawater circulation is entrained by subduction of the tectonic plates. If the temperature of the plate is high enough, a rare case for the modern mantle, the oceanic crust melts to form a transition magma between basalt and granite. If, in contrast, temperature is low, the fluids produced by dehydration of the oceanic crust escape and react with the underlying mantle to form magmas most commonly known as andesites. These magmas are saturated with respect to olivine and pyroxene. They precipitate large amounts of solid cumulate and the composition of residual liquids becomes more and more granitic-like. At the risk of giving a simplistic representation of this process, granitic rocks might be seen as dehydration melts. During the Archaean, they evolved as the result of the melting on the oceanic crust thanks to abundant volumes of water locked in hydrous minerals.

Two major questions still remain. First, what is the origin of the basaltic crust that undergoes this melting? If it is the *ordinary* oceanic crust, continuously created at mid-oceanic ridges, then continental crust should also continuously form, since oceanic crust appears at a quasicontinuous rate of $3 \pm 0.5 \text{ km}^2$ per year, and the transit time between ridge and subduction zone is quite short ($< 100 \text{ Ma}$). This is not unexpected since, just for the sheer mass of the mantle, its thermal inertia is enormous while oceanic crust production is intimately related to the extraction of deep thermal energy. In fact, the formation of the continental crust seems to be accelerating during intense but rather short events, separated by about 300 Ma (see Fig. 3.4, Martin 2005). The episodic character of these events of young crust formation is clearly not an artifact. Their magnitude defies any probability of preferential erosion and their spatial extent is remarkable. The bulk of the new crust forming almost all of West Africa, from the South of Morocco to Ghana and from Chad to Senegal and even to the Guyana shield was extracted from the mantle in about 100 My , around 2.1 Ga ago. The episode of crustal growth at 2.7 Ga led to the formation of more than 10% of the modern continental surface. It is important to understand the nature of the events that led to the formation of such a large volume of granitic rocks. The only geological process of sufficient magnitude is the eruption of hotspot magmas, which in a few to a few tens of millions of years, by decompression melting, are known to cover areas equivalent to France or Texas with thick basaltic flows. On the continents, these flows formed the huge traps of Ethiopia, the Deccan, Parana and Karoo. Around 10% of the ocean floor is covered with similar eruptions

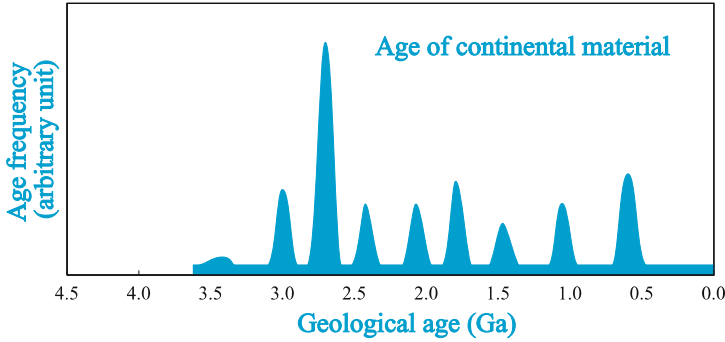


Fig. 3.4. This simplified diagram shows the episodic character of continental crust formation (see the peak at 2.7 Ga). Such a regime is difficult to reconcile with steady plate subduction. Continental crust formation during episodic events suggests a correlation with the outburst of hotspots, notably with the emplacement of oceanic plateaus

and these particularly shallow oceanic zones are known as oceanic plateaus. The basalts occurring in the Superior Province of Canada (2.7 Ga), in West of Africa (2.1 Ga) or in Arabia (600 Ma) frequently have the same geochemical characteristics as modern oceanic plateaus. The fact that the emplacement of such plateaus feeds continental growth does not change anything to the necessity that granitic melts must form along subduction zones (as they form today, for example, along the coast of Peru).

A second major point revolves around the notion of craton. Some continental zones are mobile and could be reworked by recent plate tectonic activity, while other efficiently resisted this reworking and remained virtually untouched for billions of years. West Africa was barely modified by plate tectonics for 2.1 Ga, and southwest Australia for 3.4 Ga. These stable zones are called continental cratons. Why some segments of the continental crust are mobile while others are preserved as cratons remains to be understood. The answer may tentatively be related to the origin of the subcontinental lithospheric mantle. A commonly held, yet unwarranted view, is that this part of the mantle represents remnants of the melting episode that produced the overlying crust. The lithosphere is also believed to have thickened with time because of the cooling of the upper mantle. Diamond, which is only stable at depths exceeding 150 km, helps in solving this enigma. The presence of very ancient diamonds dated at >3 Ga in the lithosphere under South Africa or Australia shows that the remarkable thickness of the subcontinental lithospheric mantle is a very ancient property. It is probable that the continental crust becomes a craton when it is emplaced on top of a low density lithosphere resistant to erosion by mantle convection. This buoyancy is the result of the accumulation of highly magnesium-rich olivine (density = $\sim 3.3 \text{ g/cm}^3$), less dense than its iron-rich counterpart (density = 4.4 g/cm^3). The peridotite fragments brought up by the kimberlitic pipes in very old cratons show this richness in Mg and confirm the higher buoyancy of

the ancient subcontinental lithospheric mantle (Mg is a refractory but fairly light element). Similar fragments coming up from under younger continents are richer in Fe and thus heavier and indicate a more unstable subcontinental lithosphere. The survival of a craton is essentially conditioned by the existence of an old, rigid and light lithosphere combined with the low density of the continental crust itself.

3.4 The Early Ages of Our Planet

Erosion and plate tectonics have destroyed the primordial crust, either oceanic or continental. The oldest segment of coherent crust on which man can walk for several kilometers is the 3.8Ga old Isua terrane of western Greenland. It is composed of peridotites, basalts, and chemically precipitated silica and iron-rich sediments (banded iron formations: BIF). All these rocks have been strongly modified by temperatures in the 400–600°C range and intense circulation of water and CO₂-rich fluids that took place when the tectonic processes assembled the terranes of this region some 3Ga ago. Such transformations are known as metamorphism. It is important to understand that, although the Isua rocks were formed in near-surface conditions (they are known as supracrustal), the metamorphism that affected them almost a Ga after their formation has preserved very little of their original features. Consequently the significance of some rocks, of potential biological interest because they contain carbon either as graphite or carbonate, remains ambiguous. Is this carbon derived from biological processes? Does it result from complex reactions between abiotic rocks and deep-originating fluids enriched in carbon dioxide or monoxide or in methane? The case of Isua carbonates that were described by the first geologists as the oldest limestone sediments is illuminating, since then it was demonstrated that they were only hydrothermal deposits, precipitated from fluids, which had passed through alternating layers of granite and peridotites (Rose et al. 1996).

Other crustal segments, of the same age as Isua (Akilia, close to the capital Nûk of Greenland, and Porpoise Cove (Nuvvuagittuq) on the east coast of the Hudson Bay), are even more modified, with metamorphic conditions reaching around 800°C. However, these modifications have not erased all traces of the original lithologies, and locally some primary geochemical characteristics can still be identified, occasionally with abundant details.

The oldest rocks known are highly deformed granites (gneiss) of Acasta far north in Canada, which are dated at 4.03Ga (Bowring and Williams 1999). They are composed of decametric boudins stretched over younger rocks that only provide a rather blurry picture of the original nature of the crust. The oldest terrestrial materials identified so far are zircon crystals (ZrSiO₄), a mineral capable of resisting alteration and having a particularly high melting temperature. They were discovered in old sandstones and conglomerates at Jack Hills in Western Australia. Unfortunately, they are included in the heart of more recent

zircons. The oldest age measured in the Jack Hills zircons is 4.404Ga (Wilde et al. 2001).

Key geochemical information can be extracted from these old minerals and rocks. The high $^{18}\text{O}/^{16}\text{O}$ ratio measured in the heart of these zircons indicates that the granitic liquids from which they crystallized came from a source already altered at low temperatures by interactions with liquid water (Mojzsis et al. 2001). So these zircons witness the presence of a continental crust just 150Ma after the formation of the solar system. The isotopic composition of neodymium in the Isua basalts shows the existence of an anomaly of the 142 isotope (Boyet et al. 2003; Caro et al. 2003). The half-life $T_{1/2}$ of the ^{142}Nd parent isotope (^{146}Sm) is 103Ma. This implies that when the basalts formed, their mantle source was still containing traces of ^{146}Sm (Fig. 3.5). The parting of the parent (^{146}Sm) and daughter (^{142}Nd) nuclides indicates that the terrestrial mantle underwent phase changes during the first 100 to 200Ma of the planet history. It is largely admitted today that this change of phase corresponded to the crystallization of a magma ocean. Enough gravitational energy was released during the accretion to melt the upper mantle. As this energy is proportional to the square of the radius of the planet, it is essentially liberated in the uppermost layers. Another argument in favor of the existence of a magma ocean is the age of the core. The extinct radioactivity ^{182}Hf - ^{182}W ($T_{1/2} = 9\text{Ma}$) indicates that core segregation was completed less than 30Ma after the isolation of the solar nebula from nucleosynthetic processes (see Chap. 2 in this volume). As already mentioned, this event itself released enough gravitational energy to warm the whole planet up to 1500°C and therefore to melt quite a significant part of it. As convection quickly brings the hottest parts of mantle back toward the surface, fusion of the upper layers seems unavoidable.

Questions can now be asked on the chemical and mineralogical nature of the primordial crusts (at this point they cannot yet be qualified as oceanic or continental). From this point of view, the size of the Earth puts it into a unique situation in the whole solar system. A peridotite lithosphere covered with basalts seems unlikely: every day, plate tectonics demonstrates that the lithosphere gets thicker and heavier as it cools down. When its thickness reaches 75 to 100km, it becomes dense enough to sink into the mantle, explaining thereby the origin of the modern plates, and today no stable buoyant lithosphere of oceanic type seems possible. A primordial lithosphere covered with important volumes of granites is unlikely as well. As we have seen, granitic melts do not form in equilibrium with mantle rocks and therefore, a hydrated basaltic crust must have predated any granitic crust.

The first geodynamic modellers (Davies 1990) rejected the idea of a magma ocean on Earth based on the argument that the continuous collapse of the crust formed at the surface of the molten layer would regularly bring up new magma, leading to the crystallization of the mantle in a few hundreds of thousands years. Nevertheless, such a model neglects the presence of water at the Earth's surface.

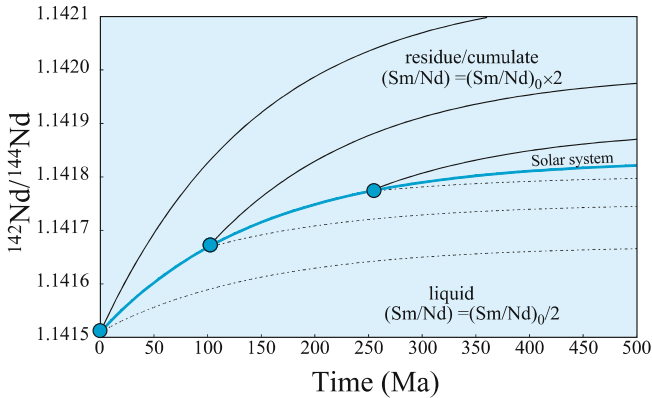


Fig. 3.5. Early fractionation of silicates in planetary mantles, in particular during the crystallization of their magma ocean, can be identified through the ^{142}Nd anomalies created by the extinct radioactivity $^{146}\text{Sm} \rightarrow ^{142}\text{Nd}$ ($T_{1/2} = 103\text{Ma}$). This diagram shows the evolution of the $^{142}\text{Nd}/^{144}\text{Nd}$ ratio in the first hundreds of Ma of the solar system evolution. The minerals (cumulates or residues) have a higher parent/daughter ratio (Sm/Nd) than that of the liquid they equilibrated with. This plot illustrates the evolution of the $^{142}\text{Nd}/^{144}\text{Nd}$ ratio assuming a twofold increase of the Sm/Nd ratio in the solid over the chondritic (solar) value and a twofold decrease in the liquid. The ultimate isotopic difference also depends on the age of Sm/Nd fractionation (*open circles*). The ^{142}Nd anomalies observed in the 3.8Ga old rocks of Isua, Greenland, require that the Earth's mantle was largely molten (magma ocean) during the first 150Ma of its history (Boyet et al. 2003; Caro et al. 2003)

The existence of the Jack Hills zircons, and therefore of granitic rocks, at $\sim 4.4\text{Ga}$ shows that water was present very early on in the history of the planet. This water probably was brought to Earth during the few tens of millions of years that Jupiter needed to expel the small, icy bodies cruising in its neighborhood. Their injection into the inner solar system and their accumulation at the surface of the rocky planets must have contributed greatly to the formation of oceans (Morbidelli et al. 2000). The important thermal absorption of water with its very large heat capacity and latent heat of vaporization is well-known. On Earth, surface water reacts immediately with dry rocks (such as basalts and peridotites) to form hydrated minerals. Several hydrated minerals are produced at temperatures intermediate between those of magmas ($\sim 1250^\circ\text{C}$) and those occurring at the surface ($0\text{--}350^\circ\text{C}$). Talc and serpentinite would dominate the upper part ($< 700^\circ\text{C}$) of the solid crust of the magma ocean. Amphiboles would become, as in the modern oceanic crust, an essential component at greater depths. These hydrated magnesium-rich minerals have lower densities (2.75 , 2.55 and $2.95\text{g}/\text{cm}^3$ for talc, serpentine, and amphibole, respectively) than the original minerals they derived from (3.3 and $3.2\text{g}/\text{cm}^3$ for olivine and pyroxene, respectively). Their presence confers to the rock a low density and helps maintain thick lithospheric buoyant rafts at the surface. The surface of a young Earth enveloped by a magma

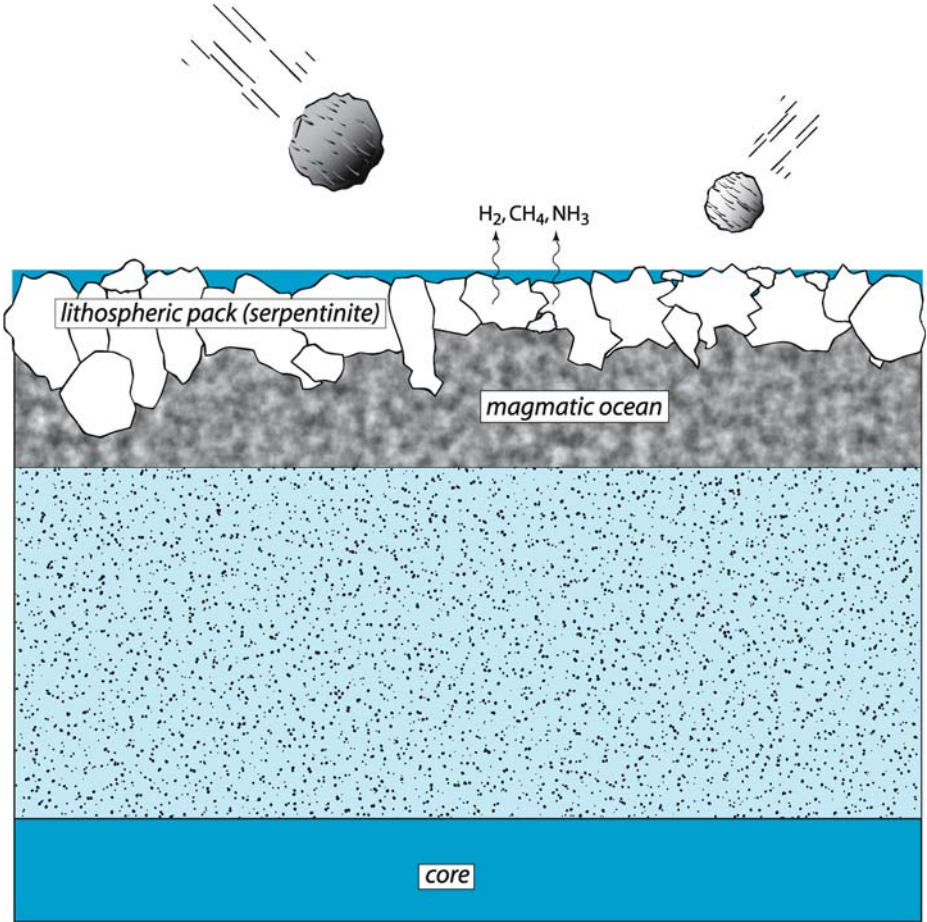


Fig. 3.6. Schematic representation of a possible primordial crust on Earth. Reaction of the partly liquid hydrosphere with the top of the magma ocean generates buoyant hydrous rocks, in which serpentinite is a dominant mineral. Olivine hydration is a strong source of hydrogen, methane, and ammonia. The steady bombardment of the early lithosphere by impacts prevented the formation of large plates and hampered the sinking of solid surface material

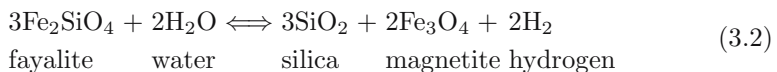
ocean and a hydrosphere had to be covered with highly hydrated magmatic rocks while being battered by a steady stream of impacts. Magmatic rock fragments most likely floated on the magma ocean, forming a dislocated lithosphere comparable to the polar ice banks (Fig. 3.6). Even if the peridotites altered in serpentinite, talc and amphibole were reworked several times during geological history, they happen to be fairly common in ancient terranes, such as those from Isua, Greenland (Fig. 3.7). Whether water was present in liquid or vapor form



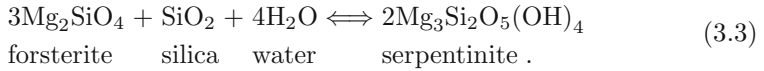
Fig. 3.7. Outcrop of hydrated peridotites (*greenish* serpentinite and *pinkish* talc) at Isua (Greenland) dated at 3.8Ga. Such material was probably particularly abundant in the primitive lithosphere

does not alter its capacity to efficiently react with basalts and form hydrous minerals.

The consequences of the presence of water for the origin of life are remarkable. Formation of hydrous minerals releases huge volumes of hydrogen that reacts with carbon dioxide to give methane, and with nitrogen to produce ammonia. Hydrogen, methane and ammonia are the essential ingredients of Millers experiment. Hydrogen and ammonia production is observed today wherever groundwater percolates through peridotites, like in Oman ophiolites (Neal and Stanger 1983). A similar phenomenon is also observed along the Mid-Atlantic Ridge, where seawater reacts with peridotites and basalts to produce substantial amounts of H_2 , CH_4 and NH_3 (Charlou et al. 1998). These processes are best illustrated by the reaction of olivine with water, that can be broken down into two subreactions, one describing the oxidation of the Fe-rich component (fayalite) of olivine

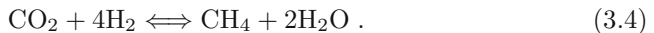


and one the hydrolysis of the Mg-rich component (forsterite) by hydrous solutions loaded with silica liberated in the previous reaction



The oxidizing hydrolysis of olivine leads to the formation of a magnetite-bearing serpentinite. At temperatures in excess of 500°C, serpentine gives way to talc ($\text{Mg}_7\text{Si}_4\text{O}_{10}(\text{OH})_2$), and then to anthophyllite ($\text{Mg}_7\text{Si}_8\text{O}_{22}(\text{OH})_2$), a mineral of the amphibole group. Since olivine is the major mineral of the upper mantle, important volumes of serpentinite and thus also of hydrogen are expected to be produced next to the surface. Free hydrogen must have reacted quite rapidly with the earliest atmosphere rich in CO_2 and N_2 , long before being able to escape the Earth's gravitational field.

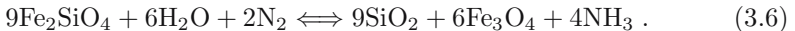
In a simplified way, the reactions involved are (Barnes and O'Neil 1969)



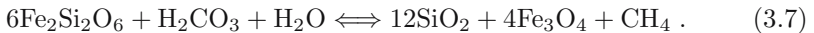
The methane-producing reaction reads



while the ammonia-producing reaction is



Other Fe and Mg-rich minerals also get involved in oxidizing hydrolysis as, for example, the ferrous component of orthopyroxene ($\text{Fe,Mg})_2\text{Si}_2\text{O}_6$, which is destroyed by a similar reaction



From the beginning, the suspicions surrounding the Miller experiment were based on the vision of a primordial oxidized atmosphere dominated by nitrogen and carbon dioxide (Kasting 1993). Sleep et al. (2004) emphasized that the huge sources of hydrogen, methane and ammonia represented by serpentinization in general and by other reactions involving other Fe^{2+} minerals, had severely been underestimated. I suggest here, that the crust overlying the magma ocean best describes the most likely environment in which these types of reactions took place. This proposition is reinforced by strong analogies existing in the modern world, notably along mid-ocean ridges. A reassessment of the conditions favorable for the origin of life based on the *old* experiment of Miller is urgently needed.

3.5 From One Planet to the Next

The chemical composition of the different rocky planets is most likely not uniform because considerable isotopic differences are observed, at least for the $^{18}\text{O}/^{16}\text{O}$

ratios of minerals, rocks and fluids. Oxygen isotope ratios are presented in the literature using the delta notation $\delta^{18}\text{O}$ in per mil defined as

$$\delta^{18}\text{O} = \left[\frac{(^{18}\text{O}/^{16}\text{O})_{\text{sample}}}{(^{18}\text{O}/^{16}\text{O})_{\text{stand}}} - 1 \right] \times 1000 \quad (3.8)$$

where the standard is usually present-day seawater. A similar notation is used for the $^{17}\text{O}/^{16}\text{O}$ ratios. It can be demonstrated that starting from a single oxygen reservoir, the relative isotopic variations of the $^{18}\text{O}/^{16}\text{O}$ ratio induced by (nearly) all natural processes are nearly exactly twice that of the $^{17}\text{O}/^{16}\text{O}$ ratio. More precisely, the thermodynamic or kinetic isotopic variations only depend on the difference between the respective masses involved and remain perfectly correlated in such a way that $\delta^{18}\text{O} = 2\delta^{17}\text{O}$. The planetary objects whose $\delta^{18}\text{O}$ and $\delta^{17}\text{O}$ values do not fall on a line with a slope of 2 reflect variations in the primary stockpiles of oxygen used by the parent bodies. It can be observed that the $\delta^{18}\text{O}$ and $\delta^{17}\text{O}$ values of the Moon and Earth fall on a single line of slope equal to 0.5 (Fig. 3.8) and thus that the oxygen of the two planets originated from a unique stockpile of material. The isotopic similarity of the oxygen supplies of both planets is a good argument against the model of a simple orbital capture of the Moon by the Earth and favors the assembly of material collected on similar orbits. The other planetary bodies from which we have samples, such as Mars (the SNC meteorites) and Vesta (achondrites) fall on different but parallel alignments, which requires that their oxygen originated from reservoirs different from the Moon–Earth reservoir. There is no current consensus about possible reasons for this heterogeneity. It has been observed that the calcium and aluminum-rich refractory inclusions, which are common in some classes of chondrites, are particularly rich in ^{16}O (Clayton 1993). The ^{16}O excesses could be inherited from pre-solar processes but how they escaped turbulent mixing during accretion is unclear. Alternatively, it could have formed inside the solar nebula by the irradiation of gas and particles during its first phases of collapse.

The distance from the Sun may also explain the compositional variability of the planets because the emitted radiation varies as the inverse of the square of the distance to the star. In the contraction (T Tauri) phase of the nebula, electromagnetic radiation (UV, X-rays) was sufficiently intense to heat up dust and force away the gases. In the inner solar system, temperature most likely exceeded 1000°C . The volatilization of certain elements by the radiating heat is easily observed in the different families of meteorites. Whereas the CI carbonaceous chondrites (e.g., Orgueil) have a very primitive composition, close to that of the condensed matter of the Sun, others (the CV and CO carbonaceous chondrites and ordinary chondrites) show depletion in volatile elements. Reactions between the condensed matter and the gas phase of the nebula are also observed. The accretion of warm material, volatilized to some degree, lead to significant chemical contrasts between planets. The Earth has a lower iron content than Mars, which itself is less Fe-rich than Mercury. Both Earth and Mars are

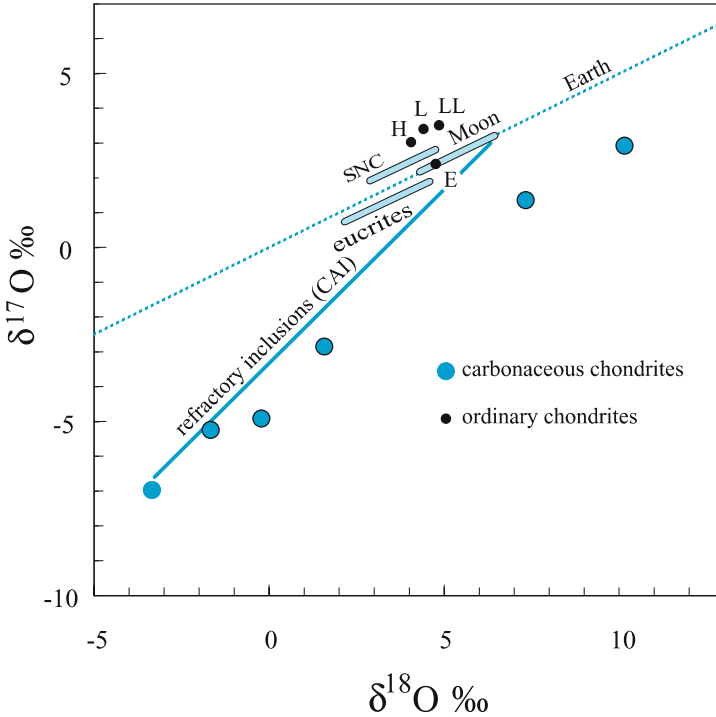


Fig. 3.8. The relative isotopic abundances of oxygen isotopes in different planetary bodies show that they originate from different initial reservoirs (Clayton 1993). Phase changes and kinetic processes modify the $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$ of the initial stockpile, yet the reaction products extracted from the same reservoir remain on a single line with a slope of 0.5 (see Martian SNC and Vesta eucrites). This is also particularly visible for terrestrial samples. The $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$ values of lunar rocks fall on the terrestrial line, which indicates that both bodies were derived from a common reservoir. The origin of the refractory inclusions (calcium-aluminum inclusions) calls for some very different processes. Various objects in the solar system have incorporated different proportions of the various oxygen components (a major element of planetary bodies). This diagram is commonly used to group meteorites and assign them to a parent body

depleted in volatile elements, such as K, Rb, Zn, Hg, compared to chondrites. The “snow line” just before Jupiter that delineates water ice accumulation on planetary bodies, also attests to the higher temperatures present in the inner solar system. Although the terrestrial planets did not form from a homogenous material, the relative proportions of the most refractory elements (Al, the rare earth, Th, U, Ti, Zr, W and some of the platinum group elements) have remained unchanged compared to the Sun and the chondrites. The temperature and the condensation of solid phases also affect iron oxidation, which results in different proportions of metallic iron Fe^0 , ferric iron Fe^{2+} and ferrous iron Fe^{3+} ; the redox states of the planets are therefore highly variable.

The parameters most affecting the chemical composition and the structure of the planets, are their size, that is, their radius, and, to a lesser degree, their density. Different radii and density lead to four consequences:

1. The mass of a rocky planet varies more or less according to the cube of its radius; planets of different size will have different masses and the resulting surface gravity will differ. As a planet size increases, it will retain more of the light elements. The escape velocity of an atom in the atmosphere is proportional to the planet radius whereas the thermal agitation velocity of the atoms in the gas is only a function of the temperature. The Earth loses hydrogen and helium-3 from its atmosphere. Mars easily loses heavier atmospheric gases such as N_2 , while the Moon and Vesta have virtually no atmosphere.
2. The acceleration of gravity (g) at the planet surface varies proportionally with the planetary radius (Fig. 3.9). The pressure inside will be affected proportionally. At the same depth, pressure in the terrestrial mantle is six times higher than in that of the Moon. The result is the existence of mineralogical contrast, which will be discussed later.
3. As mentioned above, the gravitational energy produced by planetary accretion and the formation of the core vary as the square of the radius. The upper mantle of Vesta and the Moon at the end of the accretion period were much colder than that of Earth.
4. During the first million years of the solar system, now-extinct radioactive nuclides with short half-lives, notably ^{26}Al , ^{60}Fe ($T_{1/2} = 0.7$ and 1.5Ma , respectively) were producing gamma-rays quickly absorbed by surrounding matter. Enough energy was liberated in the process to cause the melting of

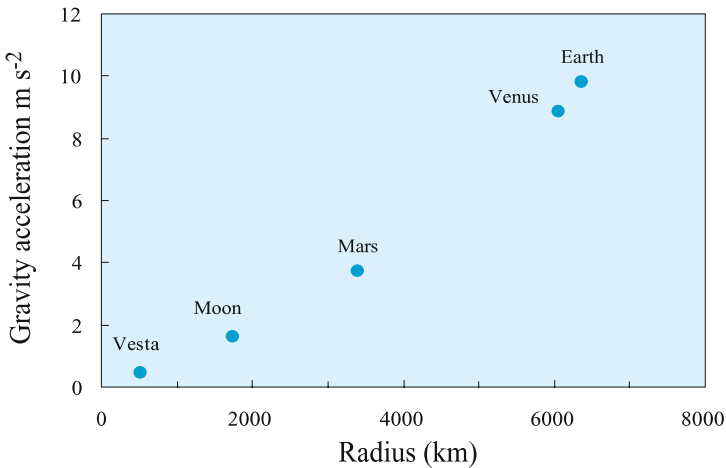


Fig. 3.9. Variation of gravity acceleration with the radius of various planets of the inner solar system

the mantle, at least as far as the radius of the planetary body was exceeding the characteristic distance for heat transport by conduction over the lifetime of the nuclides. Small asteroids, such as Vesta have completely melted and undergone a mineralogical differentiation. Others, like the parent body of the different types of chondrites, went no further than a warm up in the solid state (metamorphism).

The second consequence (2) is probably the most important factor because it controls the mineralogical composition of the mantle and crust of each planet. Table 3.1 compares the density and incompressibility of the most common minerals present in planetary mantles, and of the magma produced by their fusion. It can be seen that all the minerals are heavier than the liquids except for plagioclase. In other words, during the cooling of the magmatic, mineral saturation is reached and the dense ones sediment to the bottom, whereas buoyant plagioclase floats to the surface. On the Moon, plagioclase floating on top of the magma ocean produced the lunar highlands made of anorthosite (a rock essentially composed of plagioclase) that are without terrestrial equivalent at least in this context.

The incompressibility of the minerals and the liquid (which is the opposite of their compressibility) indicates how the solubility of minerals varies with pressure, and thus with depth in the mantle, as a function of the gravity of the planet. The incompressibility is the pressure that must be applied to these phases so that their density is multiplied by a factor e . Liquids are compressible because their structure is highly disordered and the excess pressure required to increase their density is much lower than for minerals. As pressure rises, the density of the liquid matches that of the solid, to the point that, at pressures of ~ 9 GPa, even olivine floats in basaltic melts. Plagioclase is much more compressible than the other minerals. As pressure increases, the volume change induced by the crystallization of the plagioclase becomes more important, which can only be compensated by an increase in the mineral solubility with depth. The opposite is true for pyroxene: the volume difference induced by the crystalliza-

Table 3.1. Density (ρ) and incompressibility (K) of common minerals and basaltic liquid. Note plagioclase buoyancy and compressibility.

	ρ (kg m ⁻³)	K (GPa)
olivine	3300	130
clinopyroxene	3350	115
garnet	3700	95
plagioclase	2650	65
basaltic liquid	2900	18
serpentine	2550	
talc	2600	

tion decreases when pressure increases leading to a lower solubility at greater depth. This opposite variation of solubility creates an inversion in the crystallization sequence of the two minerals at a pressure of around 0.5 GPa (Fig. 3.10). Below this limit, (arrow A, Fig. 3.10), at mid-oceanic ridge, on the Moon, or on Vesta, the cooling of basaltic magma leads to the successive precipitation of olivine-plagioclase-pyroxene, this crystallization order becomes olivine-pyroxene-plagioclase if the pressure is above 0.5 GPa (arrow B, Fig. 3.10). During the cooling of the magmatic liquid, the early crystallization of the plagioclase at lower pressure produces important volumes of plagioclase. This mineral migrates towards the surface to form oceanic gabbros (on Earth), anorthositic crust (on the Moon), and cumulative achondrite (on Vesta). At greater depths, saturation in olivine and pyroxene is reached first and these very dense minerals are deposited at the bottom of the magmatic reservoir. The lunar mantle illustrates well the accumulation of olivine and pyroxene at the bottom of the primordial magma ocean.

Let us now examine the case of the planets from which we have samples.

Vesta. It has been suggested that the parent body of the basaltic achondrites is the asteroid Vesta. Its gravity is negligible, but the presence of ^{26}Mg (Bizzaro et al. 2004), of ^{60}Fe (Shukolyukov and Lugmair 1993) and of ^{53}Cr (Shukolyukov and Lugmair 1998) in some achondrites indicates that even in the absence of gravitational energy, the extinct radioactivities led to the melting of this small planet in less than a few million years. However, it is still difficult to understand how liquid separation from the solid residue could be so efficient and therefore

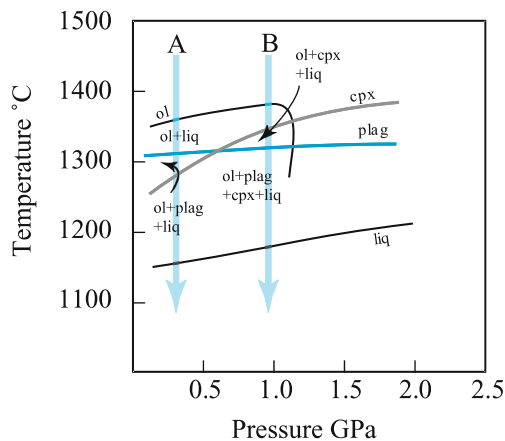


Fig. 3.10. Solubility of minerals in a basaltic melt as a function of temperature and pressure. Pressure is proportional to depth, but also to the acceleration of the planet gravity (g). Cooling at low pressure (arrow A), on Vesta or the Moon, for example, leads to the formation of large volumes of buoyant plagioclase whereas cooling at higher pressure (arrow B), as in the Earth, favors crystallization of dense olivine and plagioclase

how volcanic eruptions could take place. The basaltic achondrites are magmatic rocks rich in plagioclase and pyroxene likely representative of the crust of the planet. Cumulates of plagioclase, probably a few hundred million years younger, are also known. Other meteorites (in particular diogenites and brachinites) are part of the same group and could be representative of the mantle of this planet. All these meteorites are linked by the similar isotopic composition of their initial stock of oxygen. The fractionation of the plagioclase during the magmatic process is visible in the rock mineralogy but also in the unusual variations of in the abundances of Eu and Sr, which are commonly concentrated in this mineral.

The Moon. By size, the Moon is the closest parent to Vesta, and has gravity equal to 1/6 that of Earth. The Apollo 11 astronauts brought back samples of lunar basalts that are characterized by a major depletion in europium (Eu) compared to the other rare-earth elements (Fig. 3.11). Six months after Armstrong set foot on the Moon, Gast and Hubbard (1970) wrote: “we infer that cumulates rich in plagioclase should occur on the lunar surface,” and further down their text, predicted that the highlands were a plausible site for these plagioclase-rich rocks. The Apollo 14 and 15 missions landed close to these topographic heights and confirmed this prediction. Anorthosites are present, their age is ancient and they contain the expected Eu excess (Fig. 3.11). The recent orbital missions Galileo and Clementine have confirmed the ubiquitous character of these rocks

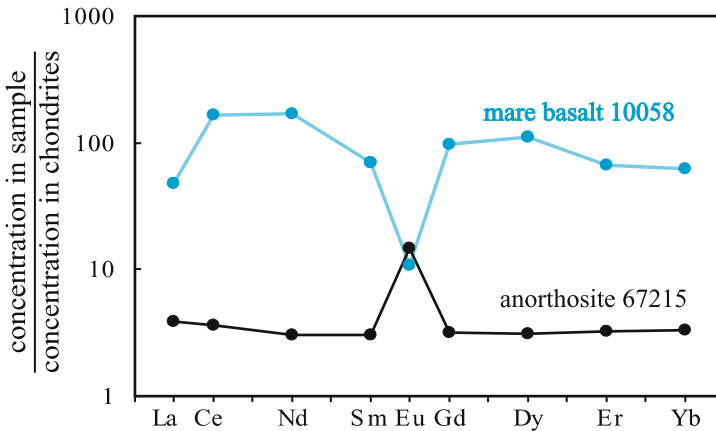


Fig. 3.11. Comparison of the concentrations of rare-earth elements (normalized to chondrites) in an Apollo 11 mare basalt (Gast and Hubbard 1970) and in an Apollo 16 anorthosite (Norman et al. 2003). The positive europium anomaly in the anorthosite sample shows that plagioclase separated from the lunar magma ocean and floated to the surface to form the highlands. The negative anomaly in the basalt sample indicates that its source corresponds to the magmatic complement of the anorthosite. Mare basalts formed by melting of olivine-pyroxene cumulates, which constitute most of the lunar mantle

at shallow depth in most lunar locations (Hawke et al. 2003). These observations demonstrated that when it was mostly molten, the planet underwent large-scale mineralogical differentiation. It is on the Moon that the concept of the magma ocean was created almost immediately after the first analyses of the Apollo 11 samples (Wood et al. 1970).

Our perception of the formation of the crust and mantle on the Moon must be related to the formation of the planet itself. The Mars-size impact scenario is currently the most popular, because it does not contradict any major geochemical or geophysical observation. The low density and iron depletion of the Moon shed doubts on the existence of a lunar core. In a low gravity environment, the production of plagioclase by the magma ocean is important and a thick layer of anorthosite is formed at the surface (lithosphere). Cooling led to the formation of a mantle most constituted of olivine and pyroxene cumulates. The residual liquid was trapped between the lithosphere and the top of the mantle. Large amounts of incompatible elements (those elements not accepted by the structure of the major minerals), such as Ti, K, the rare-earth elements, and phosphorus (KREEP), accumulated at this level. The ages of the anorthosites from the lunar mountains (Fig. 3.12) and the small anomalies of ^{142}Nd (produced by the extinct radioactivity of ^{146}Sm , see above) indicate that the magma ocean most likely persisted for more than one hundred million years. The prolonged existence of the molten layer was probably a result of the thermal blanketing effect of the

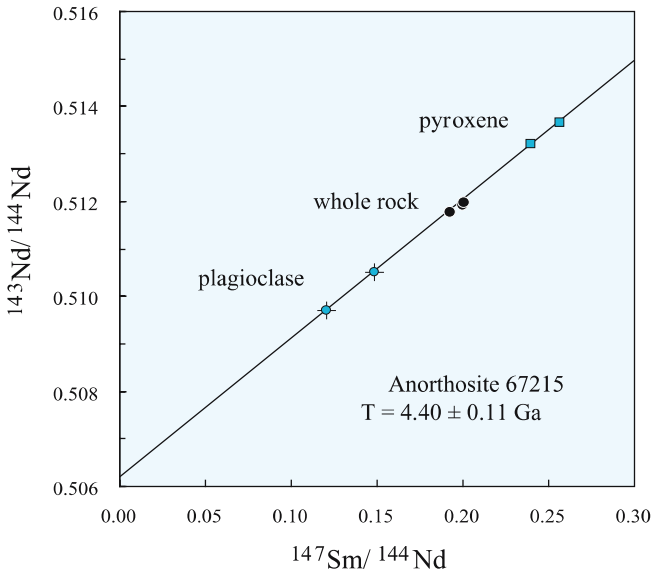


Fig. 3.12. The ^{147}Sm - ^{143}Nd isochron age obtained by Norman et al. (2003) on this anorthosite indicates that this sample is younger than planetary accretion (~ 4.56 Ga). In spite of rather large errors, this age dates the moment the rock cooled down below $\sim 700^\circ\text{C}$, strongly suggesting that the lunar magma ocean probably lasted about 100 My

anorthositic layer through which heat transfer could only occur by conduction, which is very slow.

Until around 3.2 Ga, the remelting of the lunar mantle caused by the release of radioactive heat produced mare basalt with a broad compositional variety, in particular the Ti and KREEP-rich basalts. In the meantime, the constant bombardment of the Moon had significantly modified its original surface leading to densely cratered zones. It remains difficult to assess whether the modern lunar mantle is convecting or not. The anorthosite lithosphere is stable and stagnant, and it is suspected that the extreme crystallization of the magma ocean has concentrated most of the radioactive elements at the base of the lithosphere, a situation not prone to gravitational instabilities. The analyses of the spectra collected by the various lunar orbiters have produced chemical maps at the scale of the planet (Lawrence et al. 1998). A possibility is that the concentration of U, Th, and K, strongly incompatible elements, below the anorthosite lithosphere, kept this zone molten for long periods of time (Haskin et al. 2000; Wiczoreck and Phillips 2000).

Mars. The gravity of this planet is about 1/3 of the Earth's gravity. The southern hemisphere is highly cratered and clearly very old, possibly composed of primordial crust. Younger volcanoes erupted across the Tharsis plateau, while Mount Olympus is the highest volcano in the solar system. In the absence of a calibration of the flux of impacts by dated samples, the age of the volcanic features is unclear, although they do not seem to be very recent. The northern hemisphere is covered of more recent deposits, possibly of sedimentary origin (Clifford 2001). The crustal thickness contrast inferred from the MOLA gravimetric satellite only approximately correlates with the north-south dichotomy (Zuber et al. 2000).

If no samples were brought back from Mars, there are however, today about 40 meteorites considered to be of Martian origin. They share a common oxygen isotopic signature (Fig. 3.8) and contain gas inclusions similar to those measured in the Martian atmosphere by the 1976 Viking missions. They are called SNC meteorites, according to the name of the most famous meteorites of the three defined groups: Shergotty for the basaltic meteorites, Nakhla for those dominated by pyroxenes, and Chassigny for those rich in olivine. These meteorites were highly shocked under the violence of the impact that ejected them from the planet as attested to by the amorphization of their plagioclase. The age obtained by the isotopic chronometers ^{40}K - ^{40}Ar , ^{87}Rb - ^{87}Sr , ^{147}Sm - ^{143}Nd , and U-Pb vary between 4.55 Ga (ALH84001) and 165–475 Ma (shergottites) but these young ages have recently been challenged by Pb-Pb ages of ~ 4.0 Ga obtained on plagioclase (Bouvier et al. 2005). The nakhlites and chassignites have intermediate ages around 1.3 Ga.

The recent Pathfinder and Spirit missions have analyzed the chemical and mineralogical composition of the rock surface (Gellert et al. 2004; McSween et al. 2004). They are composed of basaltic lavas with high iron concentration, which clearly appear as characteristic of Mars (and probably of Mercury) compared

to the Earth and Moon system. If the SNC originated from Mars, there are little doubts that the cratered surface of the southern hemisphere is indeed very ancient, perhaps primordial. Almost all the SNC samples display isotopic anomalies indicative of the presence of live ^{182}Hf ($T_{1/2} = 9\text{Ma}$) and ^{146}Sm ($T_{1/2} = 103\text{Ma}$) during the formation of the mantle source. This observation supports the existence of convection in the Mars mantle, and suggests that these rocks were extracted before the mantle was homogenized by convection. The implication of the young ages obtained by the long period chronometers (^{87}Rb - ^{87}Sr , ^{147}Sm - ^{143}Nd , U-Pb) is unclear. They could be interpreted as dating the emplacement of the magma, or they may be related to the consequence of the meteoritic bombardment through time or even reflect percolation of groundwater (Bouvier et al., 2005). Radiation effects in SNCs indicate that these meteorites were extracted from Mars by impacts events of well-defined ages, in particular 11 and 3Ma ago. It seems unlikely that such impact would repetitively target the youngest portions of the crust and this is a strong argument to advocate that these ages are related to the extraction of these rocks from the planet.

The internal dynamics of Mars differs greatly from that of the Moon or the Earth. Its moderate gravity did not allow for the formation of anorthositic highlands as on the Moon: in a Martian magma ocean, early crystallization of plagioclase would have been possible only down to a depth of 50km. In addition, it is likely that water was rather abundant in the early ages of Mars, thereby increasing plagioclase solubility and further reducing the depth of saturation of this mineral in magmas. Even if an anorthositic crust exists buried under thick basalt layers, meteoritic impacts have not succeeded in extracting fragments of it as meteorites.

It is likely that convection exists in the Martian mantle, but it occurs under a thick immobile lithosphere mostly composed of basalt and other magmatic rocks of similar composition. The reasons why there is apparently no plate tectonics on Mars are not well-understood but several possibilities can be put forward:

1. Water is less abundant on Mars than on Earth. The presence of an ancient northern ocean (Clifford 2001) has been recently challenged by spectroscopic observations of the Martian surface. Lack of water makes the mantle more rigid and the sliding friction between rocks during plate movement stronger. On Earth, the top of the plates is hydrated and/or covered with serpentinite and micas, which lubricate the slabs sinking into the subduction zones. The lack of water on Mars also hampers the formation of granitic rocks and continents. It has been suggested that the presence of continents, water and granites would stabilize surface constraints and would then be indispensable to the existence of plate tectonics (Campbell and Taylor 1983).
2. Because of the low gravity, the Martian lithosphere may contain a large proportion of light feldspathic rocks. Plagioclase is abundant in the SNC meteorites, even more than in the Earth's crust.

3. The scarcity of hot spots, except Tharsis, is most likely due to inefficient heat extraction from the core, where the absence of a magnetic field suggests that it is solidified.

Earth. Although it is unnecessary to repeat at this point the detailed petrological and geochemical characteristics of our planet, we will emphasize those that probably have contributed to the origin and persistence of life on this planet.

1. The presence of a partly liquid hydrosphere (ocean) on top of a mantle rich in olivine and pyroxene maintains a confined submarine reducing environment rich in hydrogen, methane, and ammonia.
2. The contact zone between solid rock and the ocean provides an environment stable on the timescale of living organisms and rich in nutriment (phosphates, nitrates) constantly re-supplied by continental weathering. It is difficult to imagine how the present biomass could survive without the permanent chemical ablation of the continental surface by erosion.
3. Plate tectonics and the continents provide a great variety of niches with various conditions favorable for the origin of life: the mid-oceanic ridges, less deep oceanic trenches, the water-sediment and the atmosphere-rock interfaces, the coastal zone and the continents (lakes, ponds). The continuous activity of the ridges renews the input of reduced gases.

The case of Mercury, for which we lack petrological and geochemical information, will not be addressed here. Venus, on the other hand, displays the traces of a recent and complete remodeling of its surface that indicates some 500Ma ago, a major change in the convection regime of the planet (Turcotte 1996).

3.6 Some Speculations

An important conclusion is therefore that the size of the Earth has allowed life to appear on it in an exclusive way. The smaller bodies with little gravity (Vesta, Moon and, perhaps, Mercury) have developed an anorthositic crust with no atmosphere and no ocean and their fate was to remain sterile. Three larger telluric planets could retain water and bring close to the surface olivine and the pyroxene. Upon hydrous alteration, these minerals released hydrogen, methane and ammonia, which are the necessary ingredients of Miller's primordial soup. Venus failed because its atmosphere was much too hot and chemically too aggressive; water seems present on this planet only as dry steam. Furthermore, its recent resurfacing event has probably destroyed any extant life. Because of its peculiar convective system, Mars has kept its original lithosphere and has not developed a plate tectonic regime. If a hydrosphere did exist in the first billion years of its evolution, water must have quickly reacted with the basement of basalt and peridotite and might have been a favorable factor contributing to the origin of life. When water eventually reached the maximum depth at which hydrous

phases are stable, the flux of reducing gas considerably declined and the loss of the primordial Martian ocean to space did the rest.

Because of its particular geodynamic, only Earth offers a setting favorable for the origin and evolution of life: a variety of niches rich in hydrogen, methane, and ammonia permitted the advent of the first proteins while continental surfaces constantly supplied organisms with fresh inputs of phosphate and nitrate and allowed them to maintain their metabolic activity.

3.7 Questions for the Future

If the origin of life and its relationship with the planetary environment have to be understood, the strongest scientific priority remains the exploration of Mars. From the point of view of a mantle geochemist, some questions are particularly relevant:

1. How deep are its mineralogical discontinuities located and what are they? Can seismology identify mantle structures indicative of a particular dynamic regime?
2. What is the age distribution at the surface?
3. Why is plate tectonics not active on Mars?
4. What is the petrologic nature of the crust; in particular, what is the importance of granitic rocks compared to basaltic lithologies?
5. Is it possible to identify geochemical markers of exchange between surface and mantle?
6. What was the distribution of the potential ancient water masses?
7. How important are the hydrated rocks and the extent of hydrous alteration at the surface? Can the remains of hydrothermal systems be identified?
8. What are the transport mechanisms of the materials produced by erosion? On Earth, priority sites are probably the study of the most ancient terranes, rocks and minerals. Among the important questions that geology and geochemistry may hope to resolve, let us list:
 9. Can extinct radioactivities identify primordial rocks?
 10. Can they confirm the existence of a magma ocean on Earth?
 11. What were the dominant materials of the first crusts: granites, basalts, serpentinite?
 12. When did plate tectonics start?
 13. When did the first emerged land appear?

Not all these questions, of course, belong to the domain of planetary science but they are among those that have a chance to contribute in a near future to astrobiology and to the understanding of the origin of life.

Acknowledgments

I would like to thank Hervé Martin, Muriel Gargaud and Brigitte Zanda for their comments on this manuscript, which have clearly improved its quality and Norm Sleep for bringing to my attention his 2004 article. I thank an anonymous translator and Philippe Claeys for the English translation.

References

General

- Albarède F. (2003) *Geochemistry: An Introduction*, 248 pp., Cambridge University Press, Cambridge
- McBride N., Gilmour I. (2003) *An Introduction to the Solar System* 418 pp., Open University Press, Cambridge
- Taylor S. R. (2001) *Solar System Evolution: A New Perspective*, 484 pp., University Press, Cambridge

Specialized

- Barnes I., O'Neil J. R. (1969) The relationships between fluids in some fresh alpine-type ultramafics and possible serpentinization, western United States. *Geol. Soc. Amer. Bull.* **80**, 1947–1960
- Bass J. D. (1995) Elasticity of minerals, glasses, and melts. In: Ahrens T. J. (ed.) *A Handbook of Physical Constants*, vol. 2: Mineral Physics and Crystallography, pp. 45–64. American Geophysical Union, Washington, DC
- Bizzarro M., Baker J. A., Haack H. (2004) Mg isotope evidence for contemporaneous formation of chondrules and refractory inclusions. *Nature* **431**, 275–2784
- Bouvier A., Blichert-Toft J., Vervoort J. D., Albarède F. (2005) The age of SNC meteorites and the antiquity of the Martian surface. *Earth Planet. Sci. Lett.* **240**, 221–233
- Bowring S. A., Williams I. S. (1999) Priscoan (4.00 ± 4.03 Ga) orthogneisses from north-western Canada. *Contrib. Mineral. Petrol.* **134**, 3–16
- Boyet M., Blichert-Toft J., Rosing M., Storey M., Telouk P., and Albarède F. (2003) ^{142}Nd evidence for early Earth differentiation. *Earth Planet. Sci. Lett.* **214**(3–4), 427–442
- Campbell I. H., Taylor S. R. (1983) No water, no granites—no oceans, no continents. *Geophys. Res. Lett.* **10**, 1061–1064
- Caro C., Bourdon B., Birck J. L., Moorbath S. (2003) ^{146}Sm – ^{142}Nd evidence from Isua metamorphosed sediments for early differentiation of the Earth's mantle. *Nature* **423**, 428–432
- Charlou J.-L., Fouquet Y., Bougault H., Donval J. P., Etoubleau J., Baptiste J.-P., Dapigny A., Appriou P., Rona P.A. (1998) Intense CH_4 plumes generated by serpentinization of ultramafic rocks at the intersection of the $15^\circ 20' \text{N}$ fracture zone and the Mid-Atlantic Ridge. *Geochim. Cosmochim. Acta* **62**, 2323–2333
- Clayton R. N. (1993) Oxygen isotopes in meteorites. *Ann. Rev. Earth Planet. Sci. Lett.* **21**, 115–149

- Clifford S. M. (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**, 40–79
- Conrad C. P., Hager B. H. (1999) The thermal evolution of an Earth with strong subduction zones. *Geophys. Res. Lett.* **26**, 3041–3044
- Davies G. F. (1990) Heat and mass transport in the early Earth. In: Newsom H. E., Jones J. H. (eds.) *Origin of the Earth*, pp. 175–194. University Press, Oxford
- Gast P. W., Hubbard N. J. (1970) Abundance of alkali metals, alkaline and rare earths, and strontium-87/strontium-86 ratios in lunar samples. *Science* **167**, 485–487
- Gellert R., Rieder R., Anderson R. C., Bruckner J., Clark B. C., Dreibus G., Economou T., Klingelhofer G., Lugmair G. W., Ming D. W., Squyres S. W., d’Uston C., Wanke H., Yen A., Zipfel J. (2004) Chemistry of rocks and soils in Gusev crater from the alpha particle X-ray spectrometer. *Science* **305**(5685), 829–832
- Haskin L. A., Gillis J. J., Korotev R. L., Jolliff B. L. (2000) The materials from the lunar Procellarum KREEP Terrane: a synthesis of data from geomorphological mapping, remote sensing, and sample analysis. *J. Geophys. Res.* **105**, 20,4003–20,415
- Hawke B. R., Peterson C. A., Blewett D. T., Busse D. B. J., Lucey P. G., Taylor G. J., Spudis P. D. (2003) Distribution and modes of occurrence of lunar anorthosite. *J. Geophys. Res.* **108**, 1–4
- Kasting J. F. (1993) Earth’s early atmosphere. *Nature* **259**, 920–925
- Lawrence D. J., Feldman W. C., Barraclough B. L., Binder A. B., Elphic R. C., Maurice S., Thomsen D. R. (1998) Global elemental maps of the Moon: the Lunar Prospector gamma-ray spectrometer. *Science* **281**, 1484–1489
- Lawrence D. J., Feldman W. C., Barraclough B. L., Elphic R. C., Prettyman T. H., Maurice S., Binder A. B., Miller M. C. (2000) Thorium abundances on the lunar surface. *J. Geophys. Res.* **105**, 20,307–20,331
- Lugmair G. W., Shukolyukov A. (1998) Early solar system timescales according to ^{53}Mn - ^{53}Cr systematics. *Geochim. Cosmochim. Acta* **62**(16), 2863–2886
- Martin H. (2005) Genesis and evolution of the primitive Earth continental crust. In: Martin H., Reisse J. (eds.) *Lectures in Astrobiology*. Springer, Berlin Heidelberg New York, pp. 113–163
- McKay D. S., Gibson E. K., Thomas-Keprta K. L., Vali H., Romanek C. S., Clemett S. J., Chillier X. D. F., Maechling C. R., Zare R. N. (1996) Search for past life on Mars: possible relic biogenic activity in Martian meteorite ALH84001. *Science* **273**(5277), 924–930
- McSween H. Y., Arvidson R. E., Bell J. F., Blaney D., Cabrol N. A., Christensen P. R., Clark B. C., Crisp J. A., Crumpler L. S., Des Marais D. J., Farmer J. D., Gellert R., Ghosh A., Gorevan S., Graff T., Grant J., Haskin L. A., Herkenhoff K. E., Johnson J. R., Jolliff B. L., Klingelhofer G., Knudson A. T., McLennan S., Milam K. A., Moersch J. E., Morris R. V., Rieder R., Ruff S. W., de Souza P. A., Squyres S. W., Wanke H., Wang A., Wyatt M. B., Yen A., Zipfel J. (2004) Basaltic rocks analyzed by the Spirit Rover in Gusev Crater. *Science* **305**(5685), 842–845
- Mojzsis S. J., Harrison T. M., Pidgeon R. T. (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., Lunine J. I., Petit J. M., Robert F., Valsecchi G. B., Cyr K. E. (2000) Source regions and timescales for the delivery of water to the Earth. *Meteor. Planet. Sci.* **35**, 1309–1320

- Neal C., Stanger G. (1983) Hydrogen generation from mantle source rocks in Oman. *Earth Plan. Sci. Lett.* **66**, 315–320
- Norman M. D., Borg L. E., Nyquist L. E., Bogard D. D. (2003) Chronology, geochemistry, and petrology of a ferroan noritic anorthosite clast from Descartes breccia 67215: clues to the age, origin, structure, and impact history of the lunar crust. *Meteor. Planet. Sci.* **38**, 645–661
- Rose N. M., Rosing M. T., Bridgwater D. (1996) The origin of metacarbonate rocks in the early Archaean Isua supracrustal belt, West Greenland. *Am. J. Sci.* **296**, 1004–1044
- Shukolyukov A., Lugmair G. W. (1993) Live iron-60 in the early solar system. *Science* **259**, 1138–1142
- Sleep N. H., Meibom A., Fridriksson T., Coleman R. G., Bird D. K. (2004). H₂-rich fluids from serpentinization: Geochemical and biotic implications. *Proc. Nat. Acad. Sci.* **101**, 12818–12823
- Sotin, C. (2001) Modèles géophysiques de l'évolution dynamique de la Terre au cours du premier milliard d'années. In: Gargaud M., Despois D., Parisot J.-P. (eds.), *L'environnement de la Terre Primitive*, pp. 237–262. Presses Universitaires de Bordeaux, Bordeaux
- Turcotte D. L. (1996) Magellan and comparative planetology. *J. Geophys. Res.* **101**, 4565–4573
- Wieczorek M. A., Phillips R. J. (2000) The Procellarum KREEP terrane: Implications for mare volcanism and lunar evolution. *J. Geophys. Res.* **105**, 20,417–20,430
- Wilde S. A., Valley J. W., Peck W. H., Graham C. M. (2001) Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* **409**, 175–178
- Wood J. A., Dickey J. S., Marvin U. B., Powell B. N. (1970) Lunar anorthosites and a geophysical model of the Moon. *Proc. Apollo 11 Lunar Sci. Conf.* **1**, 965–988
- Zuber M. T., Solomon S. C., Phillips R. J., Smith D. E., Tyler G. L., Aharonson O., Balmino G., Banerdt W. B., Head J. W., Johnson C. L., Lemoine F. G., McGovern P. J., Neumann G. A., Rowlands D. D., Zhong S. (2000) Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. *Science* **287**(5459), 1788–1793

4 Water and Climates on Mars

François Forget

4.1 Introduction

The particular importance of Mars results primarily from the existence of its atmosphere, and the potential that climatic conditions on Mars were at sometime suitable for liquid water on its surface, and thus for life. Mars' present-day climate system is complex, highly variable and poorly understood. Mars, however, is cold and dry, and its atmosphere is so thin that the presence of liquid water, never detected, is unlikely on the surface. Nevertheless, the planet was probably different in the past. The surface of Mars is characterized by multiple geological evidence that suggest that liquid water existed at and near the Martian surface at various time in its history. Two concepts of past climate (or paleoclimate) must be distinguished for Mars. On the one hand, it appears that the climate was sometimes different from what it is today throughout most of its history and until quite recently on the geological timescale (a few millions ago, or even a few thousand years ago) because of the oscillations of Mars orbit and rotation parameters. On the other hand, the observation of the geology and mineralogy of the oldest surface on Mars (dating back to more than three billion years ago) provide evidence that the Martian climate was then completely different, with abundant liquid water on the surface, probably because of a thicker atmosphere or a higher geothermal flux.

The purpose of this chapter is to briefly review what we know about the climate and water on the red planet. It is written for students or scientists who may not know anything about Mars and its climate. In the first part, I'll describe the current Martian climate system and our current understanding of the inventory of water reservoirs on present-day Mars. In the second part, I will address the recent variations of Mars' climate. Last, I will present why it is believed that early Martian climate was so much different from what it is today.

4.2 Mars' Present-Day Climate

Mars is a small planet about half the diameter of the Earth and about one and a half time more distant from the sun. The atmospheric pressures at the surface range from 1mbar to 14mbar depending on location and season (compared to

1013mbar on average at sea level on Earth). The dominant gas is CO_2 . In spite of these differences, the Martian climate system is similar to the Earth's climate system in many aspects. The two planets rotate with almost the same rate and a similar obliquity. The length of the day is thus almost the same (24h and 40 minutes on Mars) and the seasonal cycle is comparable. In such conditions, the general circulation is controlled by similar processes: on both planets, the Hadley circulation (the process that generates trade winds) is important at low latitudes, whereas "baroclinic" planetary waves (a succession of low and high-pressure zones) dominate the weather system at mid-latitudes.

A "Hypercontinental" Climate

During a nice summer afternoon in the tropics, gentle conditions can be observed on Mars with surface temperatures sometimes reaching 20°C . This does not last, however. The following night the temperature may drop by several tens of degrees and bitterly cold conditions down to -100°C can be expected until the next morning. In fact, the Martian soil, usually dry and fine-grained, cannot store much heat. Its "thermal inertia" is typically much lower than on Earth and its oceans. Since the atmosphere is very thin and cannot store much heat either, Mars' climate is characterized by very large diurnal variations. For similar reasons, the seasonal cycle is much stronger than on our planet, and the climate can be defined as being hyper-continental. Near the solstice, for instance, Mars like Earth is tilted so that the summer pole is oriented toward the sun. On both planets the region that receives the most solar energy throughout a day is then the summer pole. On our planet, the thermal inertia of the icy Arctic Ocean as well as the reflection of a large part of the solar radiation by the ice prevents the month of July at the North Pole to be warmer than, say, in Death Valley. This is not true on Mars where the maximum daily averaged temperatures are found in the high latitudes of the summer hemisphere around the solstice. The mean surface temperature then monotonically decreases from the summer pole to the winter high latitudes where temperature drops to the freezing point of CO_2 , the main constituent of the atmosphere (see below). This hemispherical asymmetry is observed most of the time except around the equinoxes where a more Earth-like regime is observed with cold poles and warmer areas around the equator. In every season, these temperature gradients drive the circulation of the atmosphere.

To the first order, the Martian meteorology can thus be compared with what one would expect on a cold, dry desert-like Earth. However, just like the climate system on Earth is not controlled only by temperature gradients and the atmospheric circulation because of the presence of oceans and water clouds, several phenomena make the climate system more complex on Mars.

First as much as 30% of the carbon dioxide atmosphere condenses every winter at high latitudes to form polar caps, inducing surface pressure variations all over the planet (CO_2 cycle). Second, a highly variable amount of suspended

dust modifies the radiative properties of the atmosphere, with sometime global dust storms totally shrouding the planet (dust cycle). Last, a peculiar hydrological cycle, characterized by water vapor transported by the atmosphere and the formation of clouds, haze and frost (water cycle), occurs on Mars.

4.2.1 The CO₂ Cycle and the Seasonal Polar Caps

The Martian polar caps are the most apparent aspects of the Martian seasonal cycle.

From Earth, they appear like white bright features waxing and waning over a Mars year. They can reach latitudes well below 50°. Analogous to the Earth, most observers assumed that the Martian polar caps were composed of water frost, until the first space probes to Mars. Following the Mariner 4 mission, Leighton and Murray (1966) used a simple thermal model to reveal the processes which control the Martian seasonal polar caps: during the fall and winter seasons at high latitudes, the local surface and atmospheric temperatures become cold enough (140 to 150K depending on the altitude) to reach the frost point of CO₂. CO₂ condenses and forms carbonic ice deposits on the surface, which can reach several tens of centimeters thick.

Although it is thought that most of the carbonic ice directly condenses on the surface, a fraction of it also condenses in the atmosphere, forming clouds and snowfall that have been detected by the Mars Global Surveyor laser altimeter MOLA (Pettengill and Ford 2000; Tobie et al. 2003) and indirectly through their infrared radiative properties (Forget et al. 1995; Titus et al. 2001).

During the spring and summer seasons in a given hemisphere, the seasonal CO₂ cap sublimates back into the atmosphere and the polar caps progressively retreat to the poles.

In the northern hemisphere, the CO₂ cover completely disappears by the beginning of summer, exposing an underlying water ice cap (see below). In the southern hemisphere, the cap recession is asymmetric and strongly varies with longitude. Furthermore, a small residual cap made of *perennial* carbonic ice survives each year through the summer. Why such a difference between the two hemispheres in spite of the fact that both poles receive the same amount of annual insolation? One can show that it is theoretically unlikely that both poles can keep a permanent CO₂ ice cap because one cap will be favored at the expense of the other. Currently, the southern polar cap is favored because of its averaged albedo, which is higher than its northern counterpart (Paige and Ingersoll 1985). By reflecting a larger amount of solar radiation, it absorbs less energy and sublimates significantly less. The difference in albedo, however, remains poorly explained. The CO₂ cycle causes the global atmospheric mass to vary by more than 30% throughout the year. Surface pressure measurements made from the two Viking probes, which landed in the northern hemisphere of the planet in 1976, clearly showed these enormous seasonal variations. Review articles on the CO₂ cycle have been published by James et al. (1992) and Forget

(1998). These articles do not include the most recent discoveries by the latest spacecraft missions however.

4.2.2 The Dust Cycle

The spacecraft mission of the 1970s has revealed that a significant amount of airborne mineral dust is always present in the atmosphere of Mars, from the surface up to an altitude of 40 km on average, and sometimes 70 km. These dust particles explain the orange color of the Martian sky as observed by the surface probes. They are about one micrometer in size and can be transported by the Martian atmosphere in spite of its tenuity. Dust can be lifted by two distinct meteorological phenomena: dust devils (convective vortices that can reach several kilometers in height and a few hundreds meters in diameter; see Thomas and Gierasch 1985; Metzger 1999), and more importantly by various kinds of dust storms. Years of monitoring by the Mars Global Surveyor shows that small dust storms occur every day on Mars (Cantor et al. 2002). Most of these events take place at the edge of the seasonal polar caps (because the thermal contrast between the carbonic ice and the bare ground induce strong thermal winds) or within the baroclinic fronts related to the low pressure zones that sweep the middle and high latitudes. Each year, a few tens of these storms last several days and reach a regional scale, with sizes of several thousand kilometers. This mostly occurs during southern spring and summer when Mars is closer to the sun on its eccentric orbit.

In some years, but not every year, one or several regional storms can expand to a global scale and shroud the entire planet for several months. This happened in 1956, 1971, 1973, 1977, and 2001. The 2001 spring global dust storm was especially well observed by the Mars Global Surveyor spacecraft (Smith et al. 2002).

What is the impact of dust on climate? During the daytime, by absorbing the solar radiation, the dust reduces the ground insolation and tends to cool the surface. At nighttime, dust infrared emission helps to keep the surface warm, so that the main effect of atmospheric dust is to reduce the amplitude of the diurnal temperature variations, without strongly affecting the surface temperatures on average. However, airborne dust has a major impact on the atmospheric temperature. At sunlit latitudes, even when the air is relatively clear, the atmosphere at 20 km is about 30 K warmer than if the atmosphere was completely dust-free. During dust storms, this heating can reach up to 80 K! Such a sensitivity has no counterpart on Earth. The presence of dust enhances the horizontal temperature gradients on the planet, and thus the atmospheric circulation. Models show that dust injected in the southern tropical atmosphere (where many regional dust storms are observed in summer) enhance the Hadley circulation and reinforce the winds in the tropics. These winds tend to lift more dust, dust reinforces the winds, etc. The development of global dust storms is thought to originate from such a positive feedback mechanism. However, explaining why global dust storms do not occur every year remains difficult (Newman et al. 2002).

4.2.3 The Water Cycle

4.2.3.1 Water on Mars Today

Is there water on Mars? Please be informed that water is almost everywhere on Mars in a solid or gaseous state.

1. First, relatively pure, white ice is directly exposed at the surface of Mars in an area of about 1000km in diameter around the north pole. The thickness of this ice layer may vary from a few millimeters at its boundary up to a few meters or even tens of meters in some locations near its center (Thomas et al. 2000).
2. This layer appears to cover a much bigger structure made of thousands of layers of ice and dust that have accumulated over more than 3000 meters in thickness (see Clifford et al. 2000). These deposits are thought to be comparatively young, preserving a record of the seasonal and climatic cycling of atmospheric CO_2 , H_2O , and dust over the past million years (Laskar et al. 2002).
3. Near the south pole, however, the small (300km), white, bright residual cap is primarily composed of carbon dioxide ice, and for a long time it was assumed that no water ice layer was present at the surface in the south polar region. However, the first observations of the imaging spectrometer Omega aboard Mars Express has revealed the presence of perennial water ice deposits at the edges of the CO_2 ice bright cap, and even in large areas tens of kilometers away from it. Most likely, the permanent CO_2 ice cap lies on an extended water ice layer (Bibring et al. 2004).

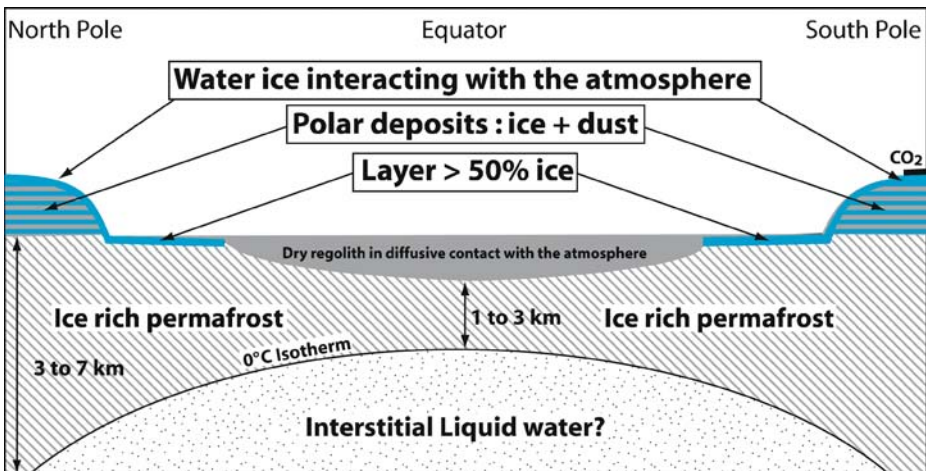


Fig. 4.1. A schematic cross section of the Martian surface and crust illustrating the various water reservoirs identified below the Martian surface (see text)

4. The analysis of the geomorphology of the south polar region, and in particular the characteristic of its topography (Smith et al. 2001) suggests that, as in the northern hemisphere, the south pole is surrounded by possibly ice-rich layered deposits which are about 1000km in diameter and 2km thick.
5. In both the northern and southern hemispheres, poleward of about 55° – 60° latitude, the Mars Odyssey Gamma-Ray Spectrometer investigation has shown that where ice is not exposed at the surface, it is still present just beneath the surface (below an ice-free layer a few centimeters thick), typically in the form of a dirty ice layer more than one meter thick and with no less than 50% of ice. What has actually been measured is the presence of large amounts of hydrogen, which can be detected through an emission of gamma rays and the capacity of hydrogen to reduce the neutron flux emitted by the surface. No other components than water ice seems likely to explain such a high amount (Feldman et al. 2002). Moreover, in these areas the unique

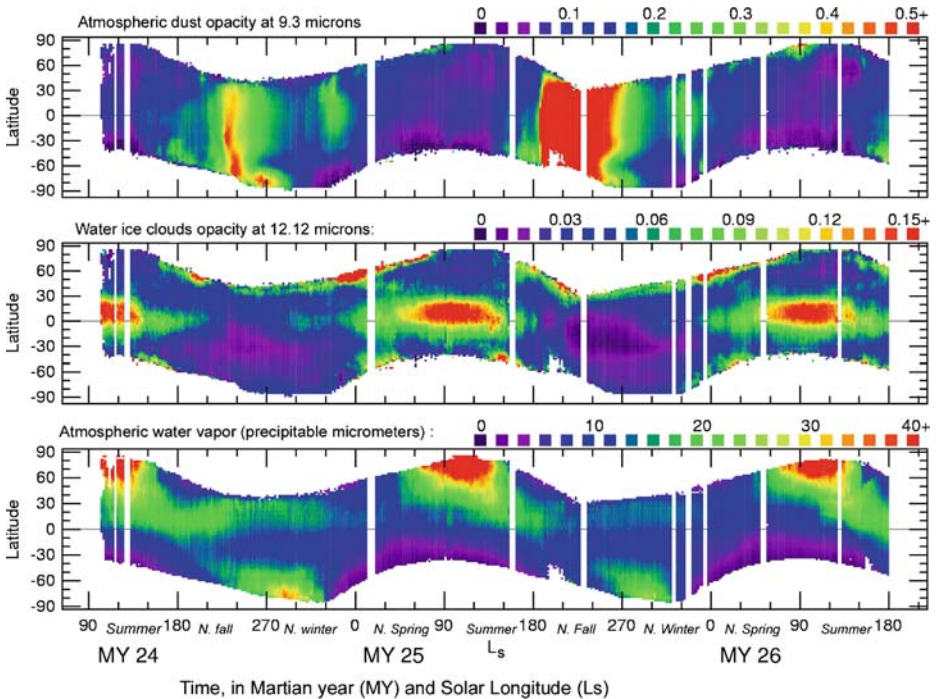
Fig. 4.2. The climate of planet Mars between 1999 and 2003, as observed by the Thermal Emission Spectrometer aboard Mars Global Surveyor (analyzed by Smith 2004). The *color plates* represent the zonally-averaged column abundance of dust, water ice and water vapor (observed in early afternoon) as a function of latitude and seasonal time. Seasons are expressed in terms of “Ls”, the aerocentric longitude of the sun ($Ls = 0^{\circ}$ at northern spring equinox, 90° at summer solstice, 180° at autumn equinox and 270° at winter solstice). Martian years are dated according to the calendar proposed by Clancy (Clancy et al. 2000) which starts on April 11, 1955 ($Ls = 0^{\circ}$). In each hemisphere the retrieval could not be performed during the coldest seasons at high latitudes (fall, winter, spring). The *top* figure shows the dust cycle. A period of relatively clear atmosphere is observed between $Ls = 0^{\circ}$ and $Ls = 160^{\circ}$, whereas the period from $Ls = 160^{\circ}$ to $Ls = 360^{\circ}$ is more dust laden. This last period is usually characterized by a few regional dust storms (which appear in red during Martian year 24). Martian year 25 was unusual, with a quasiglobal dust storm that lasted several months starting at $Ls = 180^{\circ}$ (end of June 2001). The *bottom* figure shows the water vapor cycle. The water vapor column abundance is given in units of precipitable micrometers (pr- μm). The water cycle is characterized by a maximum of water vapor above the sublimating northern polar cap around $Ls = 90^{\circ}$ (northern summer), which moves southward towards the equator in northern autumn. During southern spring and summer, the water trapped in the seasonal CO_2 ice cap is progressively released when the ground warms after the sublimation of the CO_2 . Ultimately, this creates another maximum around the south pole in early summer. The figure in the *middle* shows the seasonal cycle of water ice clouds (described by the cloud infrared opacities). On the one hand, clouds are observed at the boundaries of the polar night in both hemisphere. There, relatively warm air masses transported from lower latitudes are cooled, and water condenses. On the other hand, during the period when the atmosphere is relatively clear (northern spring and summer) and thus cold in altitude, the ascending branch of the Hadley cell transports water vapor in the vertical and creates a cloud belt between the equator and 30°N . These clouds almost completely disappear during southern spring and summer, because the dust laden perihelion atmosphere is too warm for significant amounts of water to condense

morphology of the ground is consistent with the presence of a few meters-thick ice layer coating the surface (see a synthesis in Head et al. 2003).

6. Below this dirty ice layer, and even at lower Martian latitudes, it is likely that the large pore volume of the Martian subsurface regolith might be filled by ice. Calculations of the thermodynamic stability of ground ice suggest that it can exist very close to the surface at high latitudes, but can persist only at substantial depths (several hundred meters) near the equator (Squyres et al. 1992). Observational evidence of the subsurface ice might be the presence of impact craters with distinctive lobate ejecta that are not found on the moon, and that apparently owe their morphology to the melting and vaporizing of ground ice during the impact. The size frequency distribution of these craters is consistent with the depth distribution of ice inferred from stability calculations.

4.2.3.2 The Seasonal Atmospheric Water Cycle

The cycle begins on the northern permanent polar cap, where ice is exposed at the surface. In summer, the ice is warmed by the sun and sublimates. For a few months, the north polar region becomes a source of water vapor, which is transported away within the atmosphere. During the rest of the year, most of this water ultimately comes back to the northern polar cap through various



transport mechanisms (Richardson and Wilson 2002; Montmessin et al. 2004). The cycle is closed and near equilibrium.

The amount of water involved in the seasonal water cycle is small. If one could precipitate the water content of the atmosphere on the surface, a layer thinner than a few tens of micrometers would be obtained even in the wettest regions (Fig. 4.2). In the cold Martian conditions, though, saturation is often reached. Therefore, clouds form in the atmosphere and frost can condense onto the surface. The atmospheric water vapor should also be able to diffuse into the porous subsurface. All these processes make the Martian water cycle. Is this cycle closed? Today there is one location on Mars where water condenses without ever being able to sublimate: the residual south polar cap. The permanent residual CO₂ frost is so cold (142K) that water can only exist in the solid state. It certainly serves as a cold trap for water. On average, water vapor is thus transported from the north pole to the south pole. Could that have always been the case? It is most likely that things were different 10³ to 10⁵ years ago (see below), and in particular that the cycle may have reversed in the past.

4.2.3.3 Liquid Water on Mars Today

To form liquid water – and possibly create an environment suitable for life as we know it – one not only needs water, but also temperatures above the freezing point (0°C for pure water), and pressure above 6.1mbar (Fig. 4.3). Such conditions are not uncommon on Mars, for instance in the lower plains in summer in early afternoon. They last only a few hours at most, however, and only the first few millimeters below the surface can be heated above 0°C. Because of the low vapor pressure of water in the atmosphere, any ice that may be present at the surface (in the morning, for instance) sublimates into the atmosphere well before it can melt (e.g., Haberle et al. 2001). The existence of liquid water in a metastable state in some specific circumstances has been suggested (Hecht 2002), but remains unlikely.

In such an environment, where could liquid water be found? More likely at several thousands meters below the surface, where water present in the ground pores could be heated by the geothermal flux (Squyres et al. 1992). The most optimistic scientists imagine that liquid water aquifers could exist at much shallower depths of up to a few tens of meters below the surface thanks to local geothermal activities, but, here again, this remains speculative.

4.3 A Few Million Years Ago: the Recent Martian Paleoclimates

4.3.1 Climate Changes Due to Orbital Parameter Variations

The climate that we observe on Mars at the present time is probably not the one that the planet enjoyed as recently as a few thousand years ago. There are

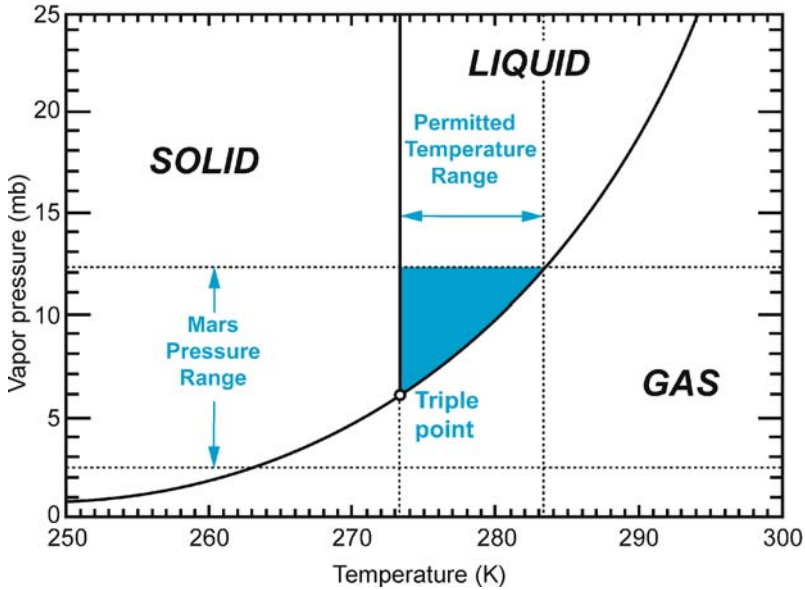


Fig. 4.3. Thermodynamic phase diagram for pure water as a function of temperature and pressure. If the air pressure (any gas) is below the water saturation pressure (*right line*), liquid water is extremely unstable (boiling). If the air pressure is above the saturation water pressure, but the partial pressure of water is below it (as often on Mars where the pressure is mostly due to CO_2), liquid water can exist, but it evaporates more or less quickly depending on the conditions just above the gas–liquid interface (the farther from the evaporation line, the more stable). In fact, it is thought that on Mars today water always sublimates into the atmosphere before it can melt (see text) (figure adapted from Haberle et al. 2001)

clear signs that the planet is in constant evolution on such timescales. Why? As on Earth, the climate on Mars depends on the planet’s orbital and rotational parameters, and in particular its obliquity (inclination of Mars axis of rotation on its orbit plane). In the case of the Earth, such oscillations are small ($\pm 1.3^\circ$ for the obliquity), but they are thought to have played a key role in the glacial and interglacial climate cycles. In the Mars case, calculations performed by Laskar’s team (Laskar and Robutel 1993; Laskar et al. 2004) have shown that Mars’ obliquity did vary widely and somewhat erratically in the past, between 0° and more than 60° (the current obliquity is 25.2°). Needless to say, these variations must have had a strong impact on Martian climate. A quantified description of these past climates is not easy because one must predict the response of a complex system that includes the atmospheric circulation and the coupled CO_2 , dust, and water cycle. Nevertheless, on the basis of theoretical considerations and with the help of numerical climate simulations, we can speculate about the major changes.

4.3.1.1 During Periods of Low Obliquity

With an obliquity below 20° , the seasonal cycle is weaker than today. The latitudinal extension of the seasonal polar caps is reduced. However, the polar regions experience a net decrease in solar heating (the sun is lower on the horizon) and thus a net cooling. It is likely that the mass of frozen atmosphere trapped at the pole (currently near the south pole) increases at the expense of the total atmosphere mass. Models also show that in such conditions the water and dust cycles are much less active. The atmosphere is thinner and clearer. Mars is then a frozen world.

4.3.1.2 During Periods of High Obliquity

Beyond 30° of obliquity, the seasonal cycle is stronger than today. For instance the CO_2 ice seasonal polar caps extend to the tropics in winter when the obliquity reaches 40° . In the polar regions, the increase of the annual mean insolation leads to a warming of the relatively deep subsurface. The CO_2 trapped in solid state in the perennial polar caps or in absorbed form in the regolith, is released until these CO_2 reservoirs are depleted (Kieffer et al. 1992). How much CO_2 can join the atmosphere-seasonal cap system is unknown, but one can imagine a Martian atmosphere thicker than today, with a more intense circulation forced by the strong seasonal cycle (Haberle et al. 2003). This atmosphere should be able to lift and maintain large amounts of airborne dust. In summer, the polar water ice reservoir is strongly heated and releases tens to hundreds of times more water than today, inducing a very active hydrological cycle (vapor, clouds, frost). Beyond obliquity of 35° to 45° , climate simulations suggest that water transported by the atmosphere from the polar regions accumulate at other latitudes (Mischna et al. 2003), most likely in the tropical regions (Levrard et al. 2004) where remains of glacier (moraines) and possibly rock covered glaciers have been observed (Head and Marchant 2003; Head et al. 2005). Climate models also suggest that, depending on the oscillations of the obliquity, water could accumulate and cover the high latitudes ($>60^\circ$) in each hemisphere with an ice layer of a few meters thick. It is likely that part of this ice is still present there today, buried below dry sediments that prevent further sublimation (Levrard et al. 2004). This would explain the Mars Odyssey observations mentioned above and the surface morphology (Head et al. 2003).

Today, it seems that a large amount of the ice that may have been transported in and out of the polar regions is back at the pole. If this scenario is confirmed, the presence of thick deposits composed of hundreds of alternatively dark and bright layers in the polar regions can probably be attributed to the past oscillations of the climate (Laskar et al. 2004).

4.3.2 Liquid Water on Mars a Few Million Years Ago

The climate variations that we have just described were probably not sufficient to allow by themselves the formation of rivers or lakes in the geologically recent past. Nevertheless, further studies are required to better evaluate if liquid water could have existed in the near subsurface for instance. On the observational side, we can mention two types of morphology, which formation probably required liquid water in the relatively recent past.

4.3.2.1 Gullies and Liquid Water

Gullies apparently formed by debris flow, including significant amounts of liquid water, have been identified on the walls of canyons, channels, impact craters (Malin and Edgett 2000a) and even dunes (Fig. 4.4). These gullies are approximately 10–30m wide, and reach several hundreds of meters long. The sharp, unweathered appearance of such small features, and the fact that debris from them overlies features that are themselves thought to be very young suggests that the gullies were formed very recently and possibly within only the last one to two million years. The process leading to their formation is still debated. Malin and Edgett (2000a) put forward a scenario involving ground water seepage from a subsurface liquid water reservoir or aquifer located a few hundred meters or less below the surface. However, the process able to maintain such a shallow aquifer at temperatures above the freezing point of water remains unclear. Geothermal heating has been invoked to melt the Martian ground ice (see also Mellon and Phillips 2001) but there is no clear association between the location of debris flows and the general distribution of recent geothermal activity. Furthermore, observations of gullies originating from isolated peaks and dune crests (Fig. 4.4, Mangold et al. 2003) question this scenario. Other scenarios usually involve past climate changes (Costard et al. 2002; Christensen 2003). For instance, Costard et al. (2002) showed that these landforms may result from the melting of liquid water in the top few meters of the Martian subsurface at high obliquity, when the mid-latitudes poleward facing slopes (where most gullies are observed) tend to accumulate water ice during fall, winter, and spring but are strongly heated above the melting point of water for several days in summer.

4.3.2.2 Recent Catastrophic Outflow

Several regions on Mars exhibit large (up to 150km across, thousands of kilometers long) channels usually referred to as outflow channels (Baker et al. 1992; Baker 2001). They appear to have been formed by catastrophic releases of large amounts of liquid water from beneath the surface (subsurface aquifer systems and/or subsurface ice melted by volcanism). Although many of the channels seem to have formed several billions years ago, some of them appear to be much

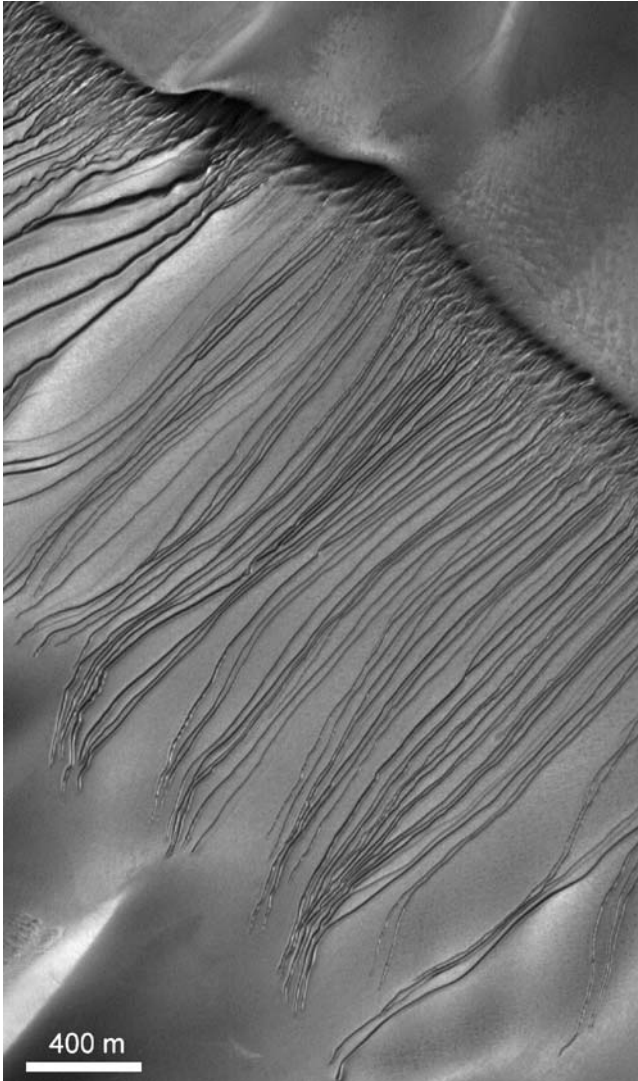


Fig. 4.4. Flow-like features (gullies) observed on a large Martian dune in the Russell crater (54°S , 347°W , Image MOC, Mars Global Surveyor, NASA/JPL/MSSS). Geomorphological characteristics like sinuosities and connections of the channels show that the formation of such gullies involved flows with 10% to 40% of liquid water (Mangold et al. 2003). These features are geologically recent: zero to a few millions years old. They were probably formed by the same processes that formed the many gullies observed on the walls of canyons, channels, and impact craters on Mars (Malin and Edgett 2000a; Costard et al. 2002). The existence of gullies originating from the crest of dunes suggests that the involvement of a subsurface aquifer is unlikely

more recent, possibly less than ten million years old (e.g., Athabasca Vallis). In fact, these catastrophic outflow channels did not have to form in a climate that is different from today's, because it is thought that when released to the surface in such large amounts, water can flow a substantial distance as a liquid prior to freezing. The presence of liquid water at the surface was thus probably short-lived.

4.4 More Than Three Billion Years Ago: the Youth of Mars

4.4.1 Evidence for Sustained Liquid Water on Early Mars

About half of the surface of Mars is extremely ancient, dating back from prior to 3.8 billions years ago, a time when the impact rate from infalling meteorites and planetesimals was much higher than today. We can estimate the age of the surface because they are densely cratered from the heavy bombardment period. This part of the Martian surface, mostly in the southern highlands, has also kept the record of a period where the environment was much different from what it became later. There is evidence that at this time, liquid water was present on the surface of Mars and could have formed rivers, lakes, and possibly an ocean. These are discussed below:

- **Valley networks.** Valleys, with characteristics very similar to terrestrial river drainage valleys, were discovered in the 1970s by the Mariner 9 orbiter. Smaller channels coalesce downhill into larger valleys, to often take the form of a dendritic network. These valleys are observed on most of the surfaces older than 3.5–3.8 billion years, but almost none have been identified on more recent terrain. This suggests that Mars enjoyed an environment suitable for the formation of these valleys more than three billion years ago (at least periodically), but that such conditions never repeated later. Was the climate different? While there is substantial debate as to whether these channels were carved by water that drained from the surface following precipitation from the atmosphere or were formed by the release of water from within the crust in a process known as sapping (cf. Carr 1996; Squyres and Kasting 1994). It is widely accepted that liquid water was involved and that it would have taken substantial periods of time to form them. This suggests that the climate must have been at least suitable for stable surface liquid water at the time. Did it rain on Mars? That is the debated issue. However, the recent exploration era is revealing more and more valley morphologies that suggest that precipitation and runoff were involved (Craddock and Howard 2002; Mangold et al. 2004).
- **Lacustrine deposits?** If Mars enjoyed a climate that allowed large amounts of liquid water to flow on its surface, it seems logical that some of this water could have accumulated in some location to form lakes, or even an ocean. This

scenario remained quite speculative until the Mars Global Surveyor camera discovered new evidence of deposits within craters that have an appearance similar to layered lacustrine deposits, deltas, sedimentary terraces and shorelines (Malin and Edgett 2000b). These deposits are thought to have been buried for long periods and were exhumed only recently (there are few impact craters superposed on the deposits, but the stratigraphic relationships suggest an old age). Well-defined layering have been identified within these deposits, supporting the idea of standing water, although windblown sediments could explain some of the observations. In some other cases however, the observations are not ambiguous. Figure 4.5, for instance, shows a delta-like structure that represent all the characteristics of alluvial deposits (see a very high definition figure at: http://www.msss.com/mars_images/moc/2003/11/13/2003.11.13a_100.jpg) (Malin and Edgett 2003).

- **Erosion rate.** A detailed analysis of the ancient terrains reveals that old impact craters larger than about 15km in diameter have a substantially degraded appearance. Crater rims, ejecta blankets, and central peaks have, in many cases, been completely removed, and the crater interiors have been filled in with debris. Craters smaller than about 15km have been erased entirely. However, craters formed in regions a few hundred million years younger are much better preserved. This observation suggests that in the distant past, before 3.8 billion years ago, the erosion rate, was a thousand times more efficient than later in the history of Mars (Jakosky et al. 2001). While this erosion could have many possible causes, it is thought that liquid water was the eroding agent (Craddock and Maxwell 1993).
- **Mineralogy.** On Earth, the composition of many terrains point to their formation in the presence of liquid water (e.g., hydrated materials). On Mars, such evidence is more seldom. The observations gathered by the thermal infrared spectrometer TES aboard Mars Global Surveyor and by the near-infrared imaging spectrometer Omega aboard Mars Express suggest that most surfaces are either of volcanic origin, or covered by dust (Christensen et al. 2001; Bibring et al. 2005). However, TES has revealed the presence of three regions with exposed surface enriched in mineral hematite, an iron oxide weathering product inferred to have been precipitated from water flowing through the crust. One of these regions, Sinus Meridiani, was selected to be explored by Opportunity, one of the two Mars Exploration Rovers that began its mission in January 2004. Opportunity did confirm the presence of hematite, and made an even more interesting discovery: it detected the presence of sulfate salts (jarosite: $\text{KFe}_3(\text{SO}_4)_2(\text{OH})_6$; kieserite: $\text{MgSO}_4 \cdot \text{H}_2\text{O}$) in the sedimentary outcrop of the area, with physical characteristics indicative of aqueous transport (Squyres et al. 2005). On this basis, the Opportunity team concluded that these rocks probably recorded episodic inundation by shallow surface water, evaporation, and desiccation. At the same time, the OMEGA actually discovered and mapped hydrated sulfates – kieserite, polyhydrated sulfates and even gypsum ($\text{CaSO}_4 \cdot 2(\text{H}_2\text{O})$) – mostly

in light-toned layered terrains located in Terra Meridiani, Valles Marineris and Margaritifer Sinus (Gendrin et al. 2005). There again, these minerals are thought to constitute direct records of the past aqueous activity on Mars.

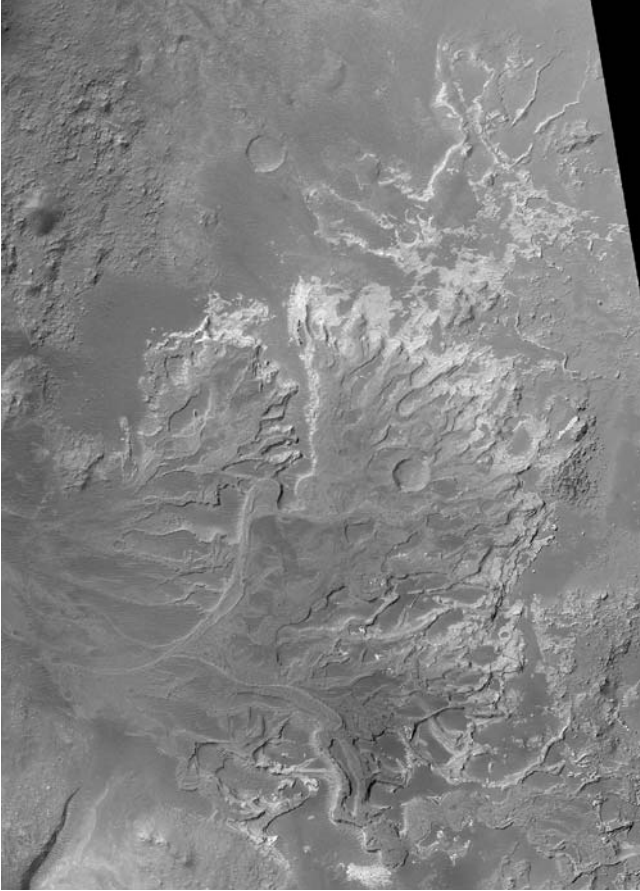


Fig. 4.5. Remnants of sedimentary alluvial fans analogous to a river delta (MOC camera, Mars Global Surveyor, NASA/JPL/MSSS). These ancient deposits have probably been covered, eroded, and exhumed throughout the long Martian geological history. The observation of such a characteristic landform provides a clear, unequivocal evidence that some valleys on Mars experienced the same type of on-going, or, persistent flow over long periods of time as rivers do on Earth, and that many of the other sedimentary rocks that are observed on Mars (but which do not exhibit such obvious delta-like characteristics) should also have been deposited in a liquid (probably water) environment (Malin and Edgett 2003)

4.4.2 The Early Mars Climate Enigma

The possible existence of relatively warm conditions suitable for surface liquid water on Mars 3.8 billions years ago is unexpected. Most experts believe that at that time, the young sun was less dense than today and its luminosity 25% lower than present time. Since Mars is about 1.5 times more distant from the sun than the Earth, the solar energy available on Mars then was only one third of what we enjoy on Earth today. In such conditions, the radiative equilibrium temperature of the planet should have been -75°C , not taking into account the atmosphere. In such conditions, what could have made the climate so warm and wet? For 20 years, this enigma has been a typical subject of collective scientific progress, with each new theory dismissing the previous one (Haberle 1998). The easiest way to warm early Mars is through the greenhouse effect of its atmosphere. The composition of the early Martian atmosphere is difficult to know for sure, but it was probably mostly composed of carbon dioxide with a surface pressure between a few hundred millibars to a few bars. Such amounts are consistent with the initial volatile inventory of a planet like Mars, and it is likely that large amounts of CO_2 and H_2O should have been released in the atmosphere by the substantial volcanism that occurred on early Mars, associated with the formation of the volcanic plains and the Tharsis bulge (Philips et al. 2001). Both CO_2 and H_2O are greenhouse gases, but their ability to warm the planet through the greenhouse effect is limited, and they may not have been able to solve the early Martian climate enigma by themselves (see review in Haberle 1998). Other greenhouse gases, such as NH_3 , CH_4 or SO_2 have been proposed to help solve the enigma, but they should have been photochemically unstable in the early Martian atmosphere and rapidly exhausted, unless produced by a surface or subsurface source (volcanism? life?). However, the presence of sulfate salts on Mars should motivate new studies involving SO_2 in the atmosphere. Another possibility is that Mars was warmed by carbonic ice clouds (Forget and Pierrehumbert 1997). These clouds tend to form in a thick CO_2 atmosphere. Models show that they can reflect a significant part of the thermal infrared radiation emitted by the surface and warm the planet through an exotic “scattering greenhouse effect”.

It is not clear how much greenhouse warming is required to explain the available geological evidence. Some of it could be explained by the large geothermal heat flux (5 to 10 times the present-time values on average) that contributed to warm the near subsurface and probably induced an intense geothermism (Squyres and Kasting 1994). One can also imagine that the large impacts that often occurred at the end of the late heavy bombardment could have played a role episodically. For instance, impacts produced global blankets of very hot ejecta that could have warmed the surface, keeping it above the freezing point of water for periods ranging from decades to millennia (Segura et al. 2002).

4.5 Conclusion

The new era of exploration of Mars that was initiated by NASA in 1997 and followed by ESA in 2003, is revolutionizing our understanding of the present and past of the planet Mars. More and more data are available, but many of these observations appear to be contradictory. In such condition, there is no consensus on what really happened on Mars, its climate evolution, the presence of liquid water, and the possibility of life.

We know that there have been multiple kinds of climates on Mars, because the planet environment strongly varied with the oscillations of the planet's orbital and rotational parameters as well as with the evolution of the content of its atmosphere. I personally would conclude that, although Mars appears to have enjoyed conditions suitable for sustained liquid water on its surface 3.7–4.2 billion years ago, it seems that afterward the Martian climates did not allow surface liquid water except during episodic events. For several billions years, Mars has probably not been very suitable for life as we know it, except maybe in the deep subsurface.

However, one can expect that new discoveries will soon, prove or disprove this opinion, bring new insight to the problems, and probably question some of the “certain facts” presented in this chapter.

References

- Baker, V.R. (2001) Water and the Martian landscape. *Nature*, 412, 228–236
- Baker, V.R., Carr, M.H., Gulick, V.C., Williams, C.R., Marley, M.S. (1992) *Channels and Valley Networks*, pp. 493–522. University of Arizona Press, Tucson, AZ
- Bibring, J.P. et al. (2004) Perennial water ice identified in the south polar cap of Mars. *Nature*, 428, 627–630
- Bibring, J.P. et al. (2005) Mars surface diversity as revealed by the OMEGA/Mars Express observations. *Science*, 307, 1576–1581
- Carr, M.H. (1996) *Water on Mars*. Oxford University Press, New York
- Chassefière, E., Leblanc F. (2004) Mars atmospheric escape and evolution; interaction with the solar wind. *Planet. Space Science*, 52, 1039–1058
- Christensen, P.R. et al. (2001) Mars Global Surveyor Thermal Emission Spectrometer experiment: investigation description and surface science results. *J. Geophys. Res.*, 106, 23 823–23 872
- Christensen, P.R. (2003) Formation of recent Martian gullies through melting of extensive water-rich snow deposits. *Nature*, 422, 45–48
- Clifford, S.M. et al. (2000) The state and future of Mars polar science and exploration. *Icarus*, 144, 210–242
- Costard, F., Forget, F., Mangold, N., Peulvast, J.P. (2002) Formation of recent Martian debris flows by melting of near-surface ground ice at high obliquity. *Science*, 295, 110–113
- Craddock, R.A., Howard, A.D. (2002) The case for rainfall on a warm, wet early Mars. *J. Geophys. Res. (Planets)*, 107, 21–1

- Craddock, R.A., Maxwell, T.A. (1993) Geomorphic evolution of the Martian highlands through ancient fluvial processes. *J. Geophys. Res.*, 98, 3453–3468
- Feldman, W.C. et al. (2002) Global distribution of neutrons from Mars: results from Mars Odyssey. *Science*, 297, 75–78
- Forget, F. (1997) Mars CO ice polar caps. In: Schmitt, B., De Bergh, C., Festou, M. (eds.) *Solar System Ices*, pp. 477–507. Kluwer Academic, Dordrecht
- Forget, F., Pierrehumbert, R.T. (1997) Warming early Mars with carbon dioxide clouds that scatter infrared radiation. *Science*, 278, 1273–1276
- Forget, F., Pollack, J.B., Hansen, G.B. (1995) Low brightness temperatures of Martian polar caps: CO clouds or low surface emissivity? *J. Geophys. Res.*, 100, 21 119–21 234
- Gendrin, A. et al. (2005) Sulfates in Martian layered terrains: the OMEGA/Mars Express view. *Science* 307, 1587–1591
- Haberle, R.M. (1998) Early Mars climate models. *J. Geophys. Res.*, 103, 28 467–28 479
- Haberle, R.M., McKay, C.P., Schaeffer, J., Cabrol, N.A., Grin, E.A., Zent, A.P., Quinn, R. (2001) On the possibility of liquid water on present-day Mars. *J. Geophys. Res.*, 106, 23 317–23 326
- Haberle, R.M., Murphy, J.R., Schaeffer, J. (2003) Orbital change experiments with a Mars general circulation model. *Icarus*, 161, 66–89
- Head, J.W., Mustard, J.F., Kreslavsky, M.A., Milliken, R.E., Marchant, D.R. (2003) Recent ice ages on Mars. *Nature*, 426, 797–802
- Hecht, M.H. (2002) Metastability of liquid water on Mars. *Icarus*, 156, 373–386
- Jakosky, B.M., Phillips, R.J. (2001) *Nature Insight Mars* special issue. **412**(6843), 237–244
- James, P.B., Kieffer, H.H., Paige, D.A. (1992) The seasonal cycle of carbon dioxide on Mars. In: Kieffer, S., Jakosky, B.M., Matthews, M. (eds.) *Mars*, pp. 934–968. University of Arizona Press, Tucson, AZ
- Kieffer, H.H., Zent, A.P. (1992) Quasiperiodic climate change on Mars. In: Kieffer, S., Jakosky, B.M., Matthews, M. (eds.) *Mars*, pp. 1180–1218. University of Arizona Press, Tucson, AZ
- Laskar, J., Robutel, P. (1993) The chaotic obliquity of the planets. *Nature*, 361, 608–612
- Laskar, J., Levrard, B., Mustard, J.F. (2002) Orbital forcing of the Martian polar layered deposits. *Nature*, 419, 375–377
- Laskar, J., Correia, A.C.M., Gastineau, M., Joutel, F., Levrard, B., Robutel, P. (2004) Long-term evolution and chaotic diffusion of the insolation quantities of Mars. *Icarus*, 170, 343–364
- Leighton, R.R., Murray, B.C. (1966) Behavior of carbon dioxide and other volatiles on Mars. *Science*, 153, 136–144
- Levrard, B., Forget, F., Montmessin, F., Laskar, J. (2004) Formation of recent Martian high-latitude ice-rich deposits by sublimation of unstable equatorial ice at low obliquity. *Nature*, 431, 1072–1075
- Malin, M.C., Edgett, K.S. (2000a) Evidence for recent groundwater seepage and surface runoff on Mars. *Science*, 288, 2330–2335
- Malin, M.C., Edgett, K.S. (2000b) Sedimentary rocks of early Mars. *Science*, 290, 1927–1937
- Malin, M.C., Edgett, K.S. (2003) Evidence for persistent flow and aqueous sedimentation on early Mars. *Science*, 302, 1931–1934

- Mangold, N., Costard, F., Forget, F. (2003) Debris flows over sand dunes on Mars: evidence for liquid water. *J. Geophys. Res. (Planets)*, 108, E4
- Mangold, N., Quantin, C., Ansan, V., Delacourt, C., Allemand, P. (2004) Evidence for precipitation on Mars from dendritic valleys in the Valles Marineris area. *Science*, 305, 78–81
- Mellon, M.T., Phillips, R.J. (2004) Recent gullies on Mars and the source of liquid water. *J. Geophys. Res. (Planets)*, 106, 23165–23180
- Melosh, H.J., Vickery, A.M. (1989) Impact erosion of the primordial atmosphere of Mars. *Nature* 338, 487–489
- Metzger, S.M., Carr, J.R., Johnson, J.R., Parker, T.J., Lemmon, M.T. (1999) Dust devil vortices seen by the Mars Pathfinder camera. *Geophys. Res. Lett.*, 26, 2781–2784
- Mischna, M.A., Richardson, M.I., Wilson, R.J., McCleese, D.J. (2003) On the orbital forcing of Martian water and CO cycles: a general circulation model study with simplified volatile schemes. *J. Geophys. Res. (Planets)*, 108, E6, 5062
- Montmessin, F., Forget, F., Rannou, P., Cabane, M., Haberle, R.M. (2004) Origin and role of water ice clouds in the Martian water cycle as inferred from a general circulation model. *J. Geophys. Res. (Planets)*, 109, E10, E10004
- Newman, C.E., Lewis, S.R., Read, P.L., Forget, F. (2002) Modeling the Martian dust cycle: 2. Multiannual radiatively active dust transport simulations. *J. Geophys. Res. (Planets)*, 107, E12, 5124
- Paige, D.A., Ingersoll, A.P. (1985) Annual heat balance of Martian polar caps: Viking observations. *Science*, 228, 1160–1168
- Pettengill, G.H., Ford, P.G. (2000) Winter clouds over the north Martian polar cap. *Geophys. Res. Lett.*, 27, 609–613
- Phillips, R.J. et al. (2001) Ancient geodynamics and global-scale hydrology on Mars. *Science*, 291, 2587–2591
- Richardson, M., Wilson, R.J. (2002) Investigation of the nature and stability of the Martian seasonal water cycle with a general circulation model. *J. Geophys. Res. (Planets)*, 107, E5, 5031
- Segura, T.L., Toon, O.B., Colaprete, A., Zahnle, K. (2002) Environmental effects of large impacts on Mars. *Science*, 298, 1977–1980
- Smith, D.E. et al. (2001) Mars Orbiter Laser Altimeter: experiment summary after the first year of global mapping of Mars. *J. Geophys. Res.*, 106, 23 689–23 722
- Smith, M.D. (2004) Interannual variability in TES atmospheric observations of Mars during 1999–2003. *Icarus*, 167, 148–165
- Smith, M.D., Conrath, B.J., Pearl, J.C., Christensen, P.R. (2002) NOTE: Thermal Emission Spectrometer observations of Martian planet-encircling dust storm 2001A. *Icarus*, 157, 259–263
- Squyres, S., Kasting, J. (1994) Early Mars, how warm and how wet? *Science*, 265, 744–749
- Squyres, S.W., Clifford, S.M., Kuz'min, R.O., Zimbelman, J.R., Costard, F.M. (1992) *Ice in the Martian Regolith*, pp. 523–554. University of Arizona Press, Tucson, AZ
- Squyres, S.W. et al. (2004) The Opportunity Rover's Athena science investigation at Meridiani Planum. *Mars Science*, 306, 1698–1703
- Thomas, P., Gierasch, P.J. (1985) Dust devils on Mars. *Science*, 230, 175–177

- Thomas, P.C., Malin, M.C., Edgett, K.S., Carr, M.H., Hartmann, W.K., Ingersoll, A.P., James, P.B. (2000) North-south geological differences between the residual polar caps on Mars. *Nature*, 404, 161–164
- Titus, T.N., Kieffer, H.H., Mullins, K.F., Christensen, P.R. (2001) TES premapping data: slab ice and snow flurries in the Martian north polar night. *J. Geophys. Res.*, 106, 23 181–23 196
- Tobie, G., Forget, F., Lott, F. (2003) Numerical simulation of the winter polar wave clouds observed by Mars Global Surveyor Mars Orbiter Laser Altimeter. *Icarus*, 164, 33–49

5 Planetary Atmospheres: From Solar System Planets to Exoplanets

Thérèse Encrenaz

5.1 What is an Atmosphere?

An atmosphere, as defined by the Greek words “atmos” (vapor) and “sphaira” (balloon), constitute, in a general manner, the gaseous envelope of a celestial body. Originally, this term has been applied to the Earth and the telluric planets, characterized by a solid surface. Today the word atmosphere is also used in the case of essentially gaseous bodies, like the giant planets or the stars (in this latter case the word refers to the external layers) or in the case of small bodies like comets.

5.1.1 Atmospheric Structure

A large variety of physical conditions can be encountered in atmospheres. All of them, however, are constrained by the hydrostatic law

$$dP = -\rho g dz \quad (5.1)$$

P being the pressure, ρ the density, g the gravity and z the altitude, which translates equilibrium between pressure and gravity. The atmospheric parameters are also constrained by the ideal gas equation

$$\rho = \mu P / RT \quad (5.2)$$

where μ is the mean molecular weight, R is the constant of ideal gases and T the temperature. As a result, the pressure follows the equation

$$dP/P = dz/H \quad (5.3)$$

where $H = RT/\mu g$ is the pressure scale height.

As the altitude increases, the pressure decreases along a law that is exponential to the first order. On Earth, at sea level, the pressure approximately decreases by a factor e ($e = 2.7$) when the altitude increases by one scale height, i.e., 8km. The scale height varies from a planet to another. It is about 10 and 14km at the surfaces of Mars and Venus, respectively, and reaches values of 20 and 40km, respectively, in the observable outer layers of Jupiter and Saturn, at

pressure levels of about 0.5bar; it is even higher (200km) in the solar photosphere. The scale height is inversely proportional to the gravity and to the mean molecular mass: the more massive the planet and the denser the atmosphere, the more concentrated the atmosphere near the surface.

In most of the atmospheres, constituents are uniformly mixed, unless they undergo condensation or chemical change. This region is called the *troposphere*. On Earth, this is the case up to the altitude of 90km, where the pressure is a few microbars. Above this level, called the homopause, each constituent decreases according to its own scale height, described by its molecular weight: this is the *heterosphere*.

How does the temperature vary along the altitude in a planetary atmosphere? In the troposphere (above the surface, in the case of terrestrial planets, or at a pressure level higher than 0.1–0.3bar in the case of the giant planets) convection takes place and the temperature decreases as the altitude increases, according to the adiabatic gradient. This is also observed in the convective regions of the Sun and stars. However, this temperature decrease is stopped at the tropopause (on Earth, this takes place at an altitude of 12km and a pressure of 220mbar; in the case of the giant planets, the tropopause level is at a pressure of about 100mbar). Above this level, the temperature increases again in the stratosphere, due to the absorption of the solar flux (mostly in the near-ultraviolet and near-infrared ranges) by various minor species (ozone on Earth, hydrocarbons and aerosols in the giant planets). In the case of Mars and Venus, there is no “stratosphere”, as this heating process does not take place; the upper atmosphere, called the *mesosphere*, is more or less isothermal, or slowly cooling as altitude increases, (except for diurnal effects), up to altitudes above 100km. On Earth, the mesosphere lays atop the stratosphere, up to the mesopause (Fig. 5.1).

In the mesosphere and above the homopause, other heating processes are involved, due to the absorption of the solar far-UV flux, which dissociates the molecules and radicals into atoms and ions. In the case of magnetized planets like the Earth, this phenomenon leads to the formation of auroras. It is also observed in the giant planets’ polar regions, where it results from the acceleration of solar wind charged particles coming from solar eruptions, which excite the ions and atoms of the upper atmosphere.

The thermal structure of a planetary atmosphere is thus the result of a balance between two energy sources, the solar energy (mostly in the UV and the near-IR spectral ranges), and the energy coming from the interior and/or from the surface, heated by the solar visible flux. In the case of the terrestrial planets, the internal heat source is negligible; for the Earth, it is only 2×10^{-4} of the solar input. In the case of all giant planets (except Uranus), the contribution from an internal energy source has to be taken into account. This balance between the two types of energy, from inside (or from the surface) and from outside, is at the origin of the thermal structure, characterized by convection in the lower levels, and radiation above the tropopause (Fig. 5.2). We could expect to find this kind

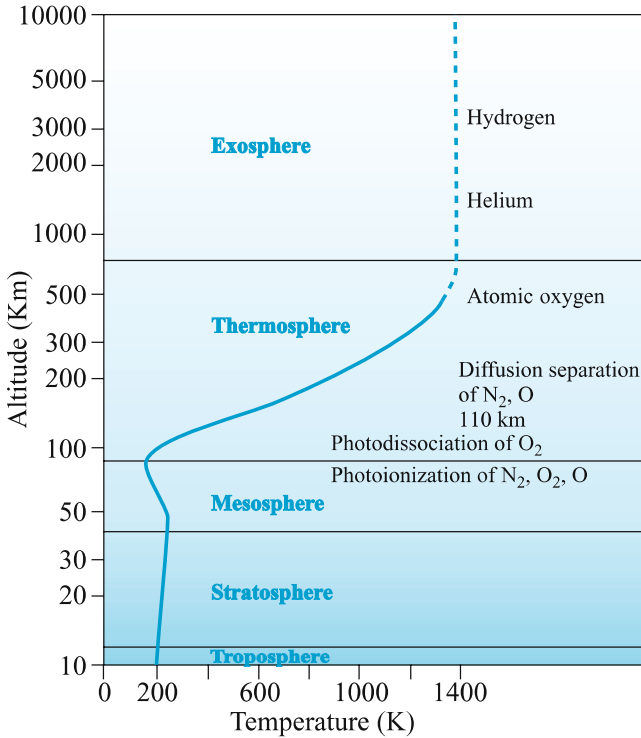


Fig. 5.1. The structure and thermal gradient of the atmosphere (after Encrenaz et al. 2004); 110 km refers to the turbopause

of structure in the newly discovered exoplanets, when these atmospheres become observable.

In the case of telluric planets, another factor is involved in the thermal structure of the lower atmosphere: the so-called *greenhouse effect*. What is this mechanism? The planetary surface is mostly heated by the solar visible radiation. In the case of terrestrial planets, the equilibrium temperature of a blackbody heated by the Sun is a few hundred degrees. As expressed by Wien's law

$$T\lambda_m = 3670 \quad (5.4)$$

where λ_m , in μm , is the wavelength of maximum radiation, and T is the temperature in K, the surface thus radiates primarily in the infrared. The wavelength of maximum emission, in the case of the Earth, is around $13\mu\text{m}$. This radiation is then very efficiently absorbed by two gases present in the terrestrial planets, carbon dioxide and water. This absorption leads to a further heating of the lower atmosphere, which, in turn, heats the surface even more, and the mechanism amplifies. On Earth, the greenhouse effect presently heats the surface by 35K. On Mars, where the atmosphere is much more tenuous than on Earth, the green-

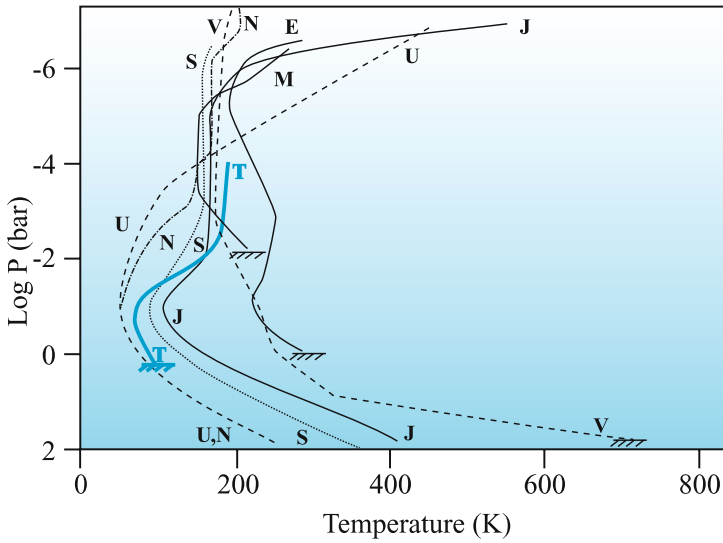


Fig. 5.2. Thermal profiles of solar system planets and satellites (after Encrenaz 2002)

house effect is only a few degrees. On Venus, in contrast, there is good evidence that the greenhouse effect is responsible for the very large difference (over 300K!) between the equilibrium and the observed temperatures. Water vapor, although a minor constituent in the Venusian atmosphere today, was probably massively present, together with carbon dioxide, in the early history of the planet. The presence of both gases probably generated a runaway greenhouse effect, with no counter-balancing mechanism. Thanks to the presence of liquid oceans, which are able to trap carbon dioxide in the form of carbonates, the greenhouse effect remained moderate on Earth.

5.1.2 Atmospheric Circulation and Cloud Structure

A main driver for atmospheric circulation is the variation of the incident solar flux, either latitudinal or seasonal. A difference of temperature between two points at a given altitude induces a motion of air masses from the warmer to the colder regions (the thermal winds). These winds are observed in the telluric planets but also on Jupiter and Saturn. Thermal winds, coupled with the planets rotation, lead to the so-called Hadley's circulation. In the case of the Earth, the warm air rises at the equator, cools down and comes down along the tropics to form two convective cells. Two other cells also form at higher latitudes in each hemisphere, between 30° and 60° , and between 60° and the pole (Fig. 5.3). Hadley cells are also observed in the other solar system planets: the faster the planet rotates, the more cells are present. In Venus, which rotates very slowly, there is one single cell in each hemisphere, while the fast rotators Jupiter and

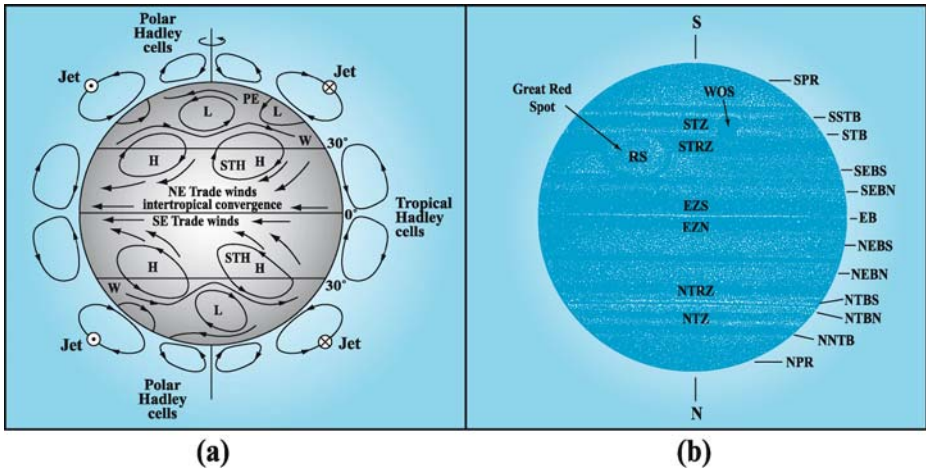


Fig. 5.3. The main features of (a) the Earth's atmospheric circulation (after Graedel and Crutzen 1993), and (b) the circulation of Jupiter (after Encrenaz et al. 2004). *SPR* = south polar region; *WOS* = white oval south; *STZ* = south temperature zone; *STRZ* = south tropical zone; *STB* = south tropical belt; *SEB* = south equatorial belt; *EZS* = equatorial zone (south). All these features have their counterpart in the northern hemisphere

Saturn exhibit a whole series of belts and zones. We note, however, that in the case of the giant planets, the presence of an internal heating source (except in Uranus) is to be taken into account for a full description of the general planetary circulation.

To understand the atmospheric motions of a planet, we need to know the nature of its condensable compounds. The situation is very simple in the case of the Earth: only one molecule, water, is able to condense. We see, however, how complex atmospheric patterns are formed, depending on the state in which water is present (liquid in ocean, rivers, rains, droplets or clouds, or solid in ice clouds, snow, ice on oceans and ground, glaciers, etc.). Typically, the water vapor contained in the warm equatorial upward motion condenses as the altitude increases and the temperature decreases. As a result, the downward motion over the tropics is mostly dried air, which explains, to the first order, the rich vegetation of the equatorial regions and the presence of deserts along the tropical belts; the reality, however, is much more complex than this simple scheme, due in particular to the interactions between the atmosphere, the oceans and the continents.

What is the cloud structure of the other planets? Venus exhibits a thick cloud cover, made of H_2SO_4 , which hides its surface permanently. In the case of Mars, two components are likely to condense, CO_2 (i.e., the main atmospheric compound) and H_2O , a very minor species in terms of relative atmospheric abundance. Because the Martian surface pressure is very low (about 6 mbars),

seasonal effects induce changes of the atmospheric total gaseous content by as much as 30%, through the seasonal transport of CO₂ from one polar cap to the other. This situation is very different from the one on Earth, which has a mean surface pressure of 1 bar; as a result, the mean atmospheric pressure, due to N₂ and O₂, remains approximately constant over the seasons. In the giant planets, mostly made of hydrogen and helium, the minor species, which are expected to condense in clouds are NH₃, NH₄SH, H₂O, with, in addition, H₂S and CH₄ in the case of Uranus and Neptune (Fig. 5.4). In all four giant planets, the photodissociation of methane leads to the formation of hydrocarbons, which partly condense (C₂H₂ and C₂H₆, in particular, are expected to condense in Uranus and Neptune). In Titan's atmosphere, dominated by N₂, the condensation of methane and the dissociation of N₂ by energetic particles, coupled with the photodissociation of methane, leads to the formation of clouds and haze presumably made of several hydrocarbons (C₂H₂, C₂H₄, C₂H₆, etc.) and nitriles (HC₃N, CH₃CN). In the very tenuous atmosphere of Triton, dominated by nitrogen, the strong obliquity of the satellite might induce seasonal effects comparable to those of Mars, with N₂ playing the role of CO₂, and CH₄ the one of Martian H₂O.

5.1.3 Atmospheric Composition

In spite of the large variety of the surface pressures and temperatures, the terrestrial planets exhibit a remarkable similarity in their atmospheric composition. In Mars and Venus, CO₂ is dominant, with a few percent of N₂ and traces of H₂O and CO. On Earth, taking into account the amounts of CO₂ and H₂O trapped in the oceans, we find also a comparable composition, with two noticeable differences: (1) water is very abundant and (2) oxygen has accumulated, as a result of the appearance of life. In the case of the giant planets, all atmospheres are dominated by much lighter elements, i.e., hydrogen and helium, with methane, ammonia and other species in reduced form (H₂O, PH₃, etc.) as minor constituents.

It is possible to understand these differences by analyzing the criteria of stability for planetary atmospheres. How can a given molecule remain in a given atmosphere over the lifetime of the solar system? To answer this question, two quantities must be evaluated and compared. The first is the thermal velocity of the molecule

$$V_{th} = (2kT/m)^{1/2} \quad (5.5)$$

where k is Boltzmann's constant, T the temperature and m the mass of the molecule. The second quantity is the escape velocity, which is the speed a body needs to reach to escape the planet's gravitational field,

$$V_{esc} = (2GM/R)^{1/2} \quad (5.6)$$

where M and R are the planet's mass and radius respectively, and G is the gravitational constant. The probability for a given molecule to escape results

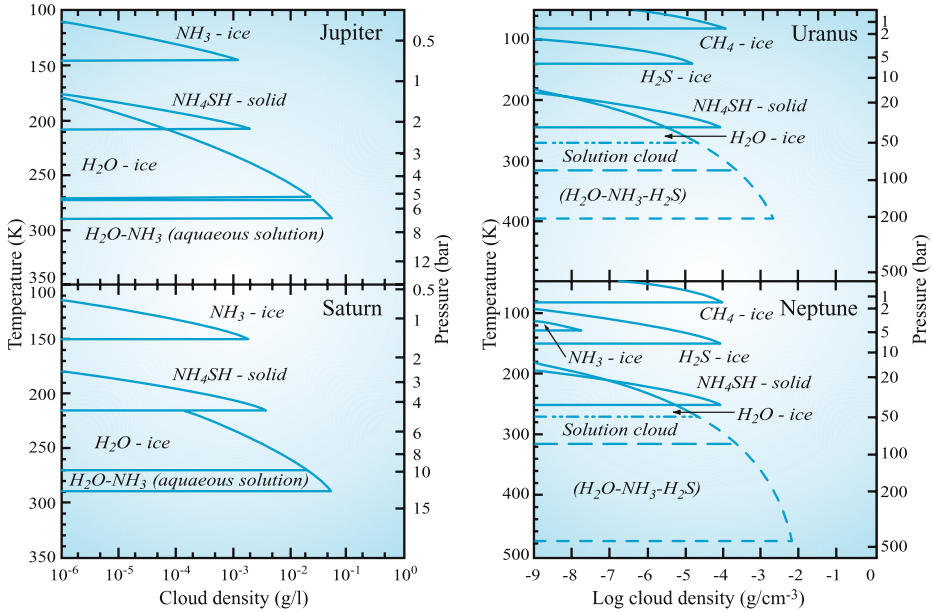


Fig. 5.4. The cloud structure of the four giant planets as inferred from thermochemical models and spectroscopic observations. This figure was constructed from models by Atreya and Romani (1985) for Jupiter and Saturn, and de Pater et al. (1991) for Uranus and Neptune (after West 1999). For Jupiter and Saturn, a simple solar composition model is assumed. Models of Uranus and Neptune imply enhanced abundance ratios, with respect to solar values. For H_2S , the enhancement factors are 10 and 30 for Uranus and Neptune, respectively. For H_2O , two cases are considered: solar ratios (*upper dashed curves*) and enhancements factors (*lower dashed curves*) by 10 (Uranus) and 30 (Neptune). It must be mentioned that, in the case of Jupiter, the cloud structure inferred from the Galileo probe, during its descent in December 1995, differed significantly from the model and showed much less cloud structure. The current interpretation is that, as a result of Jupiter's global circulation, the Galileo probe entered a region of subsidence, devoid of clouds

from a comparison of V_{th} and V_{esc} . Spitzer (1952) has shown that a molecule in a given atmosphere is stable over the solar system lifetime if V_{th} is smaller than $0.2 V_{\text{esc}}$. Molecules escape more efficiently if they are light, if their temperature is high, and if the planetary gravitational field is low. As an example, V_{esc} is equal to 11 km/s for the Earth, and 60 km/s for Jupiter. The thermal velocity of hydrogen on Earth is 2 km/s if the temperature is 300 K, 5 km/s if $T = 1500\text{K}$ (which corresponds to the ionospheric temperature). This explains why Mercury, which is small and close to the Sun, could not retain a stable atmosphere. In the case of Venus, Earth and Mars, the lightest gases like hydrogen and helium cannot remain trapped. In contrast the giant planets are massive enough to trap the lightest elements, including hydrogen and helium, over the lifetime of the

solar system. We shall see below how the formation scenario of the planets gives a global explanation for the atmospheric composition of terrestrial and giant planets.

5.1.4 Interaction with the Magnetosphere

When a planet has a magnetic field (usually generated by the dynamo effect inside a liquid or viscous core), this magnetic field interacts with the charged particles carried by the solar wind. The planetary magnetic field deviates the solar wind and creates a magnetic cavity, called the *magnetosphere*. This magnetosphere interacts with the planetary ionosphere (where atmospheric atoms and radicals are ionized) by creating electric currents, which add to the dipolar field. Charged particles are accelerated along the field lines and precipitate toward the polar regions; they excite atmospheric molecules, atoms and ions, which then emit radiation in the visible range in polar auroras. Another possible sign of the planetary magnetic field is the presence of Van Allen belts, of equatorial symmetry, which are regions where energetic particles concentrate. They emit a continuous radio emission, called synchrotron radiation, and they cross the equatorial plane at distances ranging from 1.6 to 3.7 terrestrial radii.

Solar system planets show different types of interactions with the solar wind, depending upon the existence (or not) of an atmosphere and a magnetic field. Mercury, with no atmosphere and a magnetic field, is a single case in itself. In contrast Venus, with a thick atmosphere but no magnetic field, shows an interaction comparable to that of comets. The case of Mars is probably comparable, as no permanent magnetic field has been found in the planet so far. The Earth and the four giant planets all have a dense atmosphere and a magnetic field. Jupiter's magnetosphere is the one that most resembles the terrestrial one. Finally in the case of Uranus and Neptune, the magnetic axis is tilted with respect to the rotational axis. In addition, the rotational axis of Uranus almost lies in the ecliptic plane; these peculiar geometries make each planetary magnetosphere unique.

5.2 Atmospheres of Solar System Planets

5.2.1 Formation and Evolution of Planetary Atmospheres in the Solar System

How did the solar system form? Since the statement by Copernicus in 1543 of a heliocentric system, astronomers have refined and improved formation models. We now have a global scenario of planetary formation that accounts, to the first order, for most of the observational constraints. The basic constraints are: (1) all planets (except Pluto (since August 2006, following IAU's recommendation, Pluto is no more considered as a planet)) are close to the ecliptic plane (defined

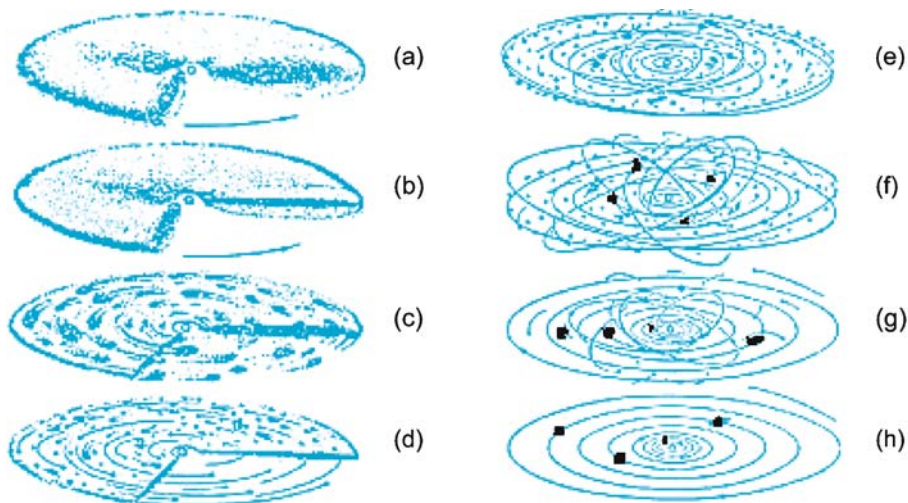


Fig. 5.5. Process of planetary formation. (a) and (b) show the formation of the disk; (c) formation of dust condensations; (d) and (e) formation of planetesimals; (f) appearance of embryos at different orbits; (g) final stages of planetary accumulation; and (h) the contemporary solar system (after Ruzmaikina 1997)

by the Earth's orbit), and (2) they all have quasicircular orbits. These two elements strongly favor the formation of a protoplanetary disk by gravitational collapse from an unstable, fast-rotating protosolar cloud. This hypothesis, first stated by Kant and Laplace in the eighteenth century, was reinforced by the observation of numerous protostellar disks around young stars over the past few decades.

According to the – now widely accepted – solar nebula model, planets formed within the protoplanetary disk. Two kinds of planetary formation models can be considered: homogeneous contraction of a local subnebula subject to instability or accretion from planetesimals made of solid particles, which grow through collisions and later gravitational attraction. In theory, the first mechanism could account for the formation of the giant gaseous planets; however, the accretion model seems to be required for the formation of the terrestrial planets, which are mostly solid. In the case of the solar system, the accretion model (Fig. 5.5) is now widely favored, not only for terrestrial planets but also for giant planets. Indeed, only this latter scenario is able to account for the chemical composition of the giant planets, as will be described below.

The accretion scenario or nucleation scenario naturally explains why two classes of planets exist in the solar system. This classification simply derives from the fact that the planets accreted from solid particles within a protosolar disk. After the collapse of the protosolar cloud into a disk, in the vicinity of the Sun ($R_h < 2\text{AU}$), the temperature must have been such that only metals and silicates could condense. This led to the formation of terres-

Table 5.1a. Orbital characteristics of the planets (from Encrenaz, 2004)

Name	Semi-major axis (AU)	Eccentricity	Inclination to the ecliptic	Sidereal period (yr)
Mercury	0.38710	0.205631	7.0048	0.2408
Venus	0.72333	0.006773	3.39947	0.6152
Earth	1.00000	0.016710	0.0000	1.0000
Mars	1.52366	0.093412	1.8506	1.8807
Jupiter	5.20336	0.048393	1.3053	11.856
Saturn	9.53707	0.054151	2.4845	29.424
Uranus	19.1913	0.047168	0.7699	83.747
Neptune	30.0690	0.008586	1.7692	163.723
Pluto	39.4817	0.248808	17.1417	248.02

terrestrial planets, relatively dense but not very massive. In contrast, beyond 2 AU (this frontier is referred as the “snow line” and was probably beyond 2 AU at the beginning of the cooling phase of the protosolar disk), most of the elements (including C, N, O, etc.) – with the exception of hydrogen and helium – were in condensed form. According to the cosmic abundances, most of the elemental mass was thus in solid form, again except H₂ and He. As ices formed with C, N and O are much more abundant than the heavier elements (Si, Mg, Fe, etc.), big solid, icy cores could then be produced with masses as high as 10–15 terrestrial masses. At smaller heliocentric distances, the solid cores reached one terrestrial mass at most. This led to a major difference in their subsequent evolution. Indeed, for large values of R_h , the massive cores were able to accrete the protosolar gas (mostly made of hydrogen and helium) through gravitational collapse (Mizuno 1980; Pollack et al. 1996). In contrast, at $R_h < 2$ AU, the gravitational field was insufficient to trap the lightest elements. As a consequence, the atmospheres of the terrestrial planets are *secondary*: they were built through outgassing and external impacts.

This simple scenario accounts in a natural way for the two classes of solar system planets, the terrestrial ones and the giant ones (Tables 5.1a and 5.1b). There is still an exception: Pluto. We know today that this remote planet is the largest representative of a new class of objects, the trans-Neptunian objects (also called the Kuiper Belt objects).

5.2.2 Terrestrial Planets and Giant Planets

5.2.2.1 Global Atmospheric Composition

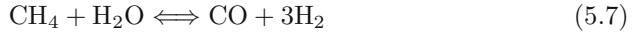
What is the expected chemical composition of planetary atmospheres? We know that the giant planets will mostly contain hydrogen and helium, and the at-

Table 5.1b. Physical properties of the planets (from Encrenaz, 2004)

Name	Sym- bol	Equatorial diameter relative to Earth	Equatorial diameter (km)	Flat- tening	Mass relative to Earth ¹	Mean density (km/s)	Surface gravity (m/s ²)	Escape velocity (day, hour, minute, second)	Sidereal rotation (day, hour, minute, second)	Inclination of the Equator to orbital plane	Principal atmospheric components
Mercury	☿	0.382	4878	0	0.055	5.44	3.78	4.25	58646 d	0°	H, He, Ne (solar wind)
Venus	♀	0.949	12104	0	0.815	5.25	8.60	10.36	243 d [Ⓔ]	2° 07'	CO ₂ (97%)
Earth	♁	1	12756	0.003353	1	5.52	9.78	11.18	23 h 56 min 04 s	23° 26'	N ₂ (78%) O ₂ (21%)
Mars	♂	0.533	6794	0.005	0.107	3.94	3.72	5.02	24 h 37 min 23 s	23° 59'	CO ₂ (95%)
Jupiter	♃	11.19	142800	0.062	317.80	1.24	24.8	59.64	9 h 50 min to 9 h 56 min	3° 04'	H, He, CH ₄ , NH ₃
Saturn	♄	9.41	120000	0.0912	95.1	0.63	10.5	35.41	10 h 14 min to 10 h 39 min	26° 44'	H, He, CH ₄ , NH ₃
Uranus	♅	3.98	50800	0.06	14.6	1.21	8.5	1.41	17 h 06 min ³	98°	H, He, CH ₄ , NH ₃
Neptune	♆	3.81	48600	0.02	17.2	1.67	10.8	23.52	15 h 48 min	29°	H, He, CH ₄ , NH ₃
Pluto	♇	0.18	2274		0.002	1.94		1.25	6 d 9 h 18 min	119.6	N ₂

¹ $m_E = 5.976 \times 10^{24}$ kg² Ⓔ indicates that the rotation is retrograde³ rotation period of the magnetic field

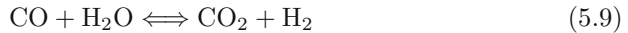
mospheres of terrestrial planets must contain heavier elements. In what form can we expect to find C, N and O in these atmospheres? Assuming thermochemical equilibrium in the protosolar nebula, we expect to find carbon in the form of CH₄ and CO, and nitrogen in the form of NH₃ and N₂. The relative abundances of these species are constrained by the following reactions:



and



evolving toward the right-hand side at high temperature and low pressure, and toward the left-hand side under opposite conditions. At large heliocentric distances, in the local subnebulae surrounding the giant protoplanets, the pressure was probably high enough, and the temperature low enough, for carbon and nitrogen to be mostly in the form of CH₄ and NH₃, respectively (Fegley et al. 1991). In contrast, in the case of the terrestrial planets, the formation of CO and N₂ must have been favored. CO was converted into CO₂ and hydrogen escaped, through the reaction



accounting, to the first order, for the global atmospheric composition of the terrestrial planets, dominated by CO₂ and N₂. In the case of the Earth, as mentioned above, H₂O is also very abundant (and it was probably abundant also in the primitive Venus), and oxygen arose later as a result of the appearance of life (Lunine 1999). Water on Earth probably came, at least partly, from asteroidal and cometary impacts.

5.2.2.2 Minor Species

In addition to the species mentioned above, planetary atmospheres also contain, in trace abundances, a large number of minor species. These minor species either came from the interior or from external interactions (photochemistry, interaction with the magnetosphere, external source).

In the case of the terrestrial planets, the internal species came through outgassing (Ar, SO₂ on Venus, primordial CH₄ on Earth, later enhanced by the appearance of life). In the case of the giant planets, they were incorporated in the planetary cores and are found in the troposphere (as PH₃, GeH₄, AsH₃ in Jupiter and Saturn; see Encrenaz et al. 1999; Encrenaz 2000 for reviews). As will be discussed below (Sect. 5.2.2.4), measuring the abundances of the tropospheric minor species in the giant planets provides important constraints upon the formation scenario of these planets.

The species coming from external interaction are mostly photodissociation products (Krasnopolsky 1986; Atreya 1986). Ozone on Earth is the best example, coming from the photodissociation of oxygen. The ozone layer makes life possible at the surface of the Earth, as it protects it from the solar UV flux, absorbed by O₃ at high altitude (Lunine 1999). In the case of the giant planets, all hydrocarbons come from the photodissociation of methane. In addition, the stratospheric H₂O and probably also CO₂ and CO come from an external oxygen source (Feuchtgruber et al. 1997), possibly connected to an interplanetary flux of micrometeorites or to a local source (rings and/or satellites). Let us mention also that some phenomena are time-dependent: as an example, the collision of comet Shoemaker-Levy 9 with Jupiter in 1994 (Noll et al. 1996) introduced new stratospheric species, some of which are still observable ten years later (CO, CS, HCN).

Tables 5.2 and 5.3 summarize the chemical compositions of planetary atmospheres.

5.2.2.3 Comparative Evolution of Terrestrial Planets

If we consider the water and the carbon dioxide trapped in the oceans, it appears that the initial atmospheric composition of Venus, Mars and the Earth are remarkably similar. However, their physical conditions are now drastically different, with surface pressures ranging from almost 100bars (Venus) down to less than 0.01bar (Mars), and surface temperatures ranging from 730K (Venus) down to 150K (minimum value on Mars).

How can we interpret these diverging histories? The answer, or at least part of it, probably lies in the initial temperature difference of the three planets. Their equilibrium temperature can be estimated from their albedo a and their heliocentric distance R_h from the following relationships, which express the balance between the absorbed solar energy and the re-emitted thermal flux:

$$T_e = (1 - a)^{1/4} \times 273 / (R_h)^{1/2} \quad \text{for a fast-rotating body (Earth and Mars)}$$

$$T_e = (1 - a)^{1/4} \times 324 / (R_h)^{1/2} \quad \text{for a slow-rotating body (Venus) .}$$

The resulting equilibrium temperatures are 364K for Venus, 250K for the Earth and 200K for Mars. The differences between these early surface temperatures were sufficient to favor the presence of water preferentially in gaseous form on Venus, liquid form on Earth and solid form on Mars. As will be discussed below, we have evidence that water was abundant in the early atmosphere of Venus, and must have contributed, as well as CO₂, to the runaway greenhouse effect of Venus. This effect did not happen on Earth because CO₂ was trapped in liquid water. In the case of Mars, two factors have to be considered: its lower temperature and its much lower mass with respect to Venus and the Earth. From the early beginning, the internal energy of Mars (mostly through the decay of radioactive nuclei) was accordingly lower. As a result, volcanism was less efficient.

Table 5.2. Atmospheric properties of terrestrial planets and outer solid bodies

Object	Surface pressure P_s (bar)	Surface temperature T_s (K)	Composition	Nature of clouds	Comment
Mercury	–	520	–	–	No stable atmosphere
Venus	90	730	CO ₂ (0.95), N ₂ (0.035) + traces of CO, H ₂ O, SO ₂	H ₂ SO ₄	Strong greenhouse effect
Earth	1	288	N ₂ (0.77), O ₂ (0.21), H ₂ (0.017), Ar (0.01), H ₂ O (<0.04) + traces of CO, O ₃ , CH ₄ , NO ₂	H ₂ O	Moderate greenhouse effect; H ₂ O in oceans
Mars	0.006	220	CO ₂ (0.95), N ₂ (0.027), Ar (0.016), O ₂ (0.0013) + traces of CO, O ₃ , H ₂ O ₂ , CH ₄	CO ₂ , H ₂ O	Weak greenhouse effect
Io	$1-4 \times 10^{-8}$	110	SO ₂ , + traces of SO, NaCl		Non-uniform atmosphere (volcanism)
Titan	1.5	95	N ₂ (0.98), CH ₄ (0.18), H ₂ (0.02) + hydrocarbons, nitriles, + traces of CO, H ₂ O, CO ₂	CH ₄ , hydrocarbons	Continuous CH ₄ out-gassing?
Triton	1.6×10^{-8}	38	N ₂ , CH ₄ (10^{-4})	Hydrocarbons	Cryo-volcanism
Pluto	10^{-8}	40	N ₂ , CH ₄ (0.01), CO (0.001)		P_s and T_s variable with R_{th}

Table 5.3. Abundances in the atmospheres of the giant planets

Species	Jupiter	Saturn	Uranus	Neptune
H ₂	1	1	1	1
HD	1.8×10^{-5}	2.3×10^{-5}	5.5×10^{-5}	6.5×10^{-5}
He	0.136	0.11–0.16	0.18	0.23
CH ₄ (trop)	2.1×10^{-3}	4.4×10^{-3}	2×10^{-2}	4×10^{-2}
CH ₄ (strat)	2.1×10^{-3}	4.4×10^{-3}	3×10^{-5} – 3×10^{-4}	7×10^{-4} (0.05–1 mb)
¹³ CH ₄ (trop)	2×10^{-5}	4×10^{-5}		
CH ₃ D (trop)	2.5×10^{-7}	3.2×10^{-7}	10^{-5}	2×10^{-5}
CH ₃ D (strat)				2.2×10^{-7}
C ₂ H ₂		3.6×10^{-6} (0.1 mb)	$2\text{--}4 \times 10^{-7}$ (0.1–0.3 mb)	1.1×10^{-7} (0.1 mb)
¹² C ¹³ CH ₂	*	2.5×10^{-7}		
C ₂ H ₆	4.0×10^{-6} (0.3–50 mb)	4×10^{-6} (< 10 mb)		1.3×10^{-6} (0.03–1.5 mb)
CH ₃ C ₂ H	*	6×10^{-10} (< 10 mb)		
C ₄ H ₂		9×10^{-11} (< 10 mb)		
C ₂ H ₄	7×10^{-9}			3×10^{-7} (< 1 μb)
C ₃ H ₈	6×10^{-7}			
C ₆ H ₆	2×10^{-9}	*		
CH ₃		$0.2\text{--}1 \times 10^{-7}$		2.9×10^{-8} (0.2 μb)
NH ₃ (trop)	2×10^{-4} (3–4 b)			
¹⁵ NH ₃	8×10^{-7}			
PH ₃ (trop)	6×10^{-7}	1.7×10^{-6}		
GeH ₄	7×10^{-10}	2×10^{-9}		
AsH ₃	3×10^{-10}	2×10^{-9}		
CO (trop)	1.5×10^{-9}	2×10^{-9}		
CO (strat)	1.5×10^{-9}	2×10^{-9}	3×10^{-8}	10^{-6}
CO ₂ (strat)	4×10^{-10} (< 10 mb)	3×10^{-10} (< 10 mb)		5×10^{-10} (< 5 mb)
H ₂ O (trop)	1.4×10^{-5} (3–5 mb)	2×10^{-7} (< 3 b)		
H ₂ O (strat)	1.5×10^{-9} (< 10 mb)	$2\text{--}20 \times 10^{-9}$ (< 0.3 mb)	$5\text{--}12 \times 10^{-9}$ (< 0.03 mb)	$1.5\text{--}3.5 \times 10^{-9}$ (< 0.6 mb)
HCN				3×10^{-10}
H ₃ ⁺	*	*	*	

* detected

In addition, because of the small size of the planet, the capture of infalling bodies was less efficient than on Venus and Earth. As a result of these two factors, the Martian atmosphere was less dense (although probably denser than today); the planet cooled down and finally became inactive. There is now evidence for the presence of water ice under the perennial Martian polar caps, as shown by the recent results of the orbiting spacecraft Mars Odyssey (Boynton et al. 2002) and Mars Express (Bibring et al. 2004).

The comparative evolution of the terrestrial planets illustrates the large variety of physicochemical conditions, which may result from relatively small differences in the early conditions of these objects. Such variety might be expected also if Earth-like exoplanets are discovered.

5.2.2.4 Outer Planets: Gaseous and Icy Giants

As discussed above, according to the current formation scenario, the giant planets formed from an icy core of about 10–15 terrestrial masses, which later accreted the surrounding protosolar nebula through gravitational collapse. If we compare the mass of this core to the total masses of the giant planets (Table 5.1), we see that this represents only 3% of Jupiter's mass, about 12% of Saturn's mass, but more than 60% of the masses of Uranus and Neptune. Jupiter and Saturn, mostly made of protosolar gas, are called the *gaseous giants*, while Uranus and Neptune, mostly composed of their icy core, are called the *icy giants*.

The presence of an icy core in the giant planets has a direct effect on their chemical composition. Indeed, if the planets had formed directly from the protosolar gas, the chemical composition should be identical to the protosolar one, and thus reflect the cosmic abundances. In contrast, the presence of an initial core of heavy elements (we refer here to all elements heavier than a mass of about 10: carbon, nitrogen, oxygen and all heavier elements) must have enriched the atmospheres of the giant planets in these elements. Assuming all the elements were equally trapped in the cores (which remains a debatable question), and assuming homogeneous mixing inside the planets after the collapse phase, it is possible to calculate the expected enrichment in heavy elements for the four giant planets. According to cosmic abundances, heavy elements account for 2% of the total mass. We infer an enrichment of a factor 3 for Jupiter, 7 for Saturn and 30–60 for Uranus and Neptune (Owen and Encrenaz 2003). The factor 3 enrichment was exactly measured by the Galileo probe for several elements (C, N, S and the noble gases), in the case of Jupiter (Owen et al. 1999; Fig. 5.6). The C/H ratio is found to be enriched by the expected factors on the four giant planets (Gautier and Owen 1986). This diagnostic has been decisive to validate the nucleation model of the giant planets. We note, however, that the O/H ratio in the deep troposphere of Jupiter is still unknown, because the low abundance measured by the Galileo probe is probably due to local meteorological processes (Atreya et al. 1999).

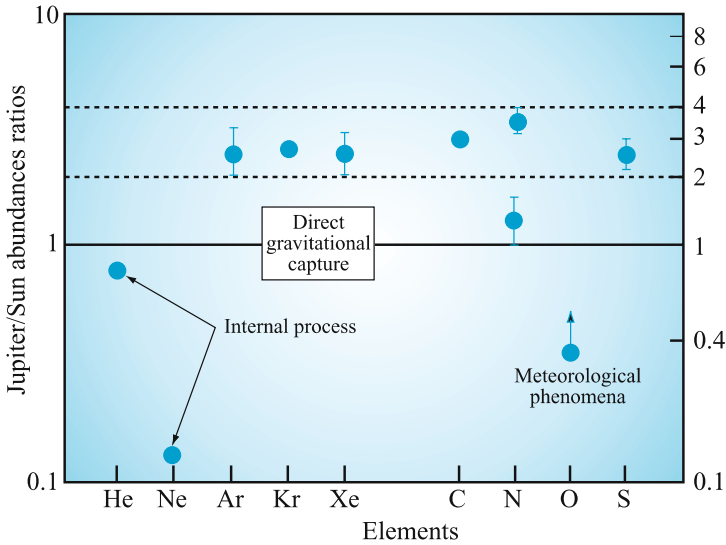


Fig. 5.6. Elemental abundances in Jupiter, measured relative to hydrogen, compared with solar values. Apart from He, Ne and O (which are believed to be depleted by internal or dynamical processes), all the elements show enrichments by a factor of about 3 (after Owen et al. 1999)

Beyond the simple classification in gaseous and icy giants, each giant planet is a world in itself. In spite of their similarities in their atmospheric compositions and structures, Jupiter and Saturn show noticeable differences (Atreya et al. 1999, 2002), in particular in their internal heat sources (larger in Saturn) and their dynamical activities (Saturn’s eddy diffusion coefficient being much stronger). Uranus and Neptune also exhibit puzzling differences: Neptune has a strong internal heat source while Uranus has none; Neptune is very active, with a high eddy diffusion coefficient, while Uranus is much more sluggish (Bergstrahl et al. 1991; Cruikshank et al. 1995). Finally, CO and HCN were found in large amounts in Neptune’s stratosphere, but not in Uranus (Rosenqvist et al. 1992; Marten et al. 1993). The peculiar geometry of Uranus (with its rotational axis close to the ecliptic plane) might be connected to these differences, which remain to be explained.

5.2.3 Atmospheres of Outer Satellites and Pluto

Like the terrestrial and giant planets, some outer satellites – as well as Pluto, when the planet is in the vicinity of perihelion – are surrounded with a stable atmosphere, which largely differs from one case to another.

In the vicinity of Jupiter, only Io, the closest Galilean satellite, has a stable atmosphere, extremely tenuous, with a surface pressure lower than 40 nanobars,

variable over time and location, and mostly composed of sulfur dioxide SO_2 (Lelouch et al. 1992). Traces of SO and NaCl have also been found. This atmosphere, constantly outgassing from the interior by active and permanent volcanism, is a unique case in the solar system. The internal heat is generated by tidal effects due to the proximity of Jupiter and the next two Galilean satellites, Europa and Ganymede. As a result, the surface of Io is constantly regenerated. Io's atmosphere is partially ionized by the solar UV flux, but also by energetic particles associated to the strong magnetic field of Jupiter. This interaction results in a plasma torus, mostly made of hydrogen, with also sulfur, oxygen and sodium ions. Io's atmosphere illustrates the case of a solid body immersed in the close vicinity of a very massive and magnetized planet.

Titan, Saturn's largest satellite, is an even more peculiar case (Coustenis and Taylor 1999). With its nitrogen-dominated atmosphere and its surface pressure of 1.5bars (Hunten et al. 1984), it presents striking similarities with the Earth, with the noticeable difference of a much colder atmosphere: the surface temperature is 91K and its minimum temperature, at the tropopause, is about 70K. Titan's atmosphere raises many questions, so far unsolved. Since N_2 is dominant, why is carbon in the form of CH_4 (at a level of about 2%) rather than CO (less than 100 parts per million)? A possible explanation could be that Titan, formed in Saturn's protosolar subnebula, was originally rich in NH_3 and CH_4 ; NH_3 could have been subsequently transformed into N_2 by photodissociation. Also, it is interesting to note that Titan has a stable atmosphere while Ganymede and Callisto, of comparable sizes, have none. The answer is probably related to a difference in internal structure, due to the temperature difference between Jupiter and Saturn's orbits. Like the giant planets, Titan's stratosphere is subject to an intense photochemistry, responsible for the formation of many hydrocarbons, some of which condense in the lower stratosphere. The originality of Titan comes from the nitrogen chemistry, induced by the destruction of N_2 by solar UV and energetic particles associated with Saturn's magnetic field. The combination of CH_4 and N_2 dissociation processes leads to the formation of nitriles, some of them (HC_3N , CH_3CN) being the first steps for the development of prebiotic chemistry (Coustenis and Lorenz 1999). Titan, which was explored in-depth by the Cassini-Huygens space mission in 2004 to 2005, is thus a target of major interest for exobiology.

Farther away from the Sun, Neptune's largest satellite Triton (Cruikshank et al. 1995) shows striking similarities with Pluto (Stern and Tholen 1997). In the 1990s, Pluto was close to perihelion, and its heliocentric distance was close to Triton's. Away from perihelion, Pluto's atmosphere is likely to freeze. Both atmospheres are dominated by N_2 with traces of CH_4 . The surface pressure is about 10 microbars on Triton, and probably comparable on Pluto (under perihelion conditions), in agreement with a gas/solid equilibrium. On Triton, the surface temperature is 38K, while it ranges between 40 and 60K on Pluto, depending on the position on the surface. Several ices have been found at the surface of both bodies: N_2 , CH_4 , CO, CO_2 on Triton and possibly H_2O on

Pluto. As we know from its orbit that Triton is a captured satellite of Neptune, the similarity between Triton and Pluto can be understood if we consider both objects as large samples of the Kuiper Belt population, which has been discovered over the past decade (Davies 2001). However, the capture mechanism of Triton is not yet understood.

5.3 Tools for Studying Planetary Atmospheres

Since the advent of the space era, the exploration of planets uses two main tools, the ground-based observations and the in-orbit or in situ measurements. It is important to note that, whereas planetary space exploration has made spectacular achievements over the past decades, ground-based observations, using the largest telescopes and the most sophisticated instruments, still provide and will in the future continue to provide unique information on planetary atmospheres and surfaces.

There are two basic experimental methods for investigating the properties of solar system objects: remote sensing analysis, through imaging and/or spectroscopy, an in situ analysis, through, in particular, mass spectrometry, or X/γ spectroscopy. Remote sensing analyses can be performed either from the ground or aboard spacecraft orbiting around planets.

5.3.1 Remote Sensing Analysis

5.3.1.1 Visible, UV and IR Imaging: Monitoring of Surface and/or Dynamical Properties

Imaging at all wavelengths is best suited for monitoring spatiotemporal variations of planetary surfaces. Sounding at different wavelengths allows probing different altitude ranges of planetary atmospheres. UV sounding probes upper atmospheres where photodissociation takes place. In the case of dense atmospheres like Venus, the Earth or giant planets, the penetration level, outside the absorption bands associated to specific stratospheric species, is limited by Rayleigh scattering. The energy of the UV solar flux is sufficient to ionize and/or dissociate molecules into radicals, atoms and/or ions. The visible solar flux probes down to a pressure level of about 1 bar in the case of dense atmospheres, and down to the surface in the tenuous atmosphere of Mars. The infrared range probes down to deeper levels in the case of Venus and the giant planets. In the spectral windows where the absorption by atmospheric constituents is low, the flux arises from deeper layers, close to the surface between the CO_2 near-IR bands in the case of Venus, and down to several bars at $4.5\text{--}5\mu\text{m}$ in the atmospheres of Jupiter and Saturn.

Imaging planetary disks at several wavelengths, knowing their atmospheric composition, thus allows for the retrieval of a three-dimensional monitoring of

their global circulation and dynamical motions. As an example, visible monitoring of Jupiter and Saturn measures meridional winds at the level of the NH_3 clouds at a pressure of about 0.5 bar. Near-infrared monitoring of Venus at $2.3\ \mu\text{m}$ allows mapping the lower layer of H_2SO_4 clouds at about 40–50 km altitude. In the case of bodies without an atmosphere, or with tenuous or transparent atmospheres, imaging planetary disks at visible and near-IR wavelengths maps the morphology and the mineralogy of the surface.

5.3.1.2 Infrared and Radio Spectroscopy: Chemical Composition, Elemental and Isotopic Abundance Ratios

Planetary spectra are characterized by two components: a solar component corresponding to the reflected part of the solar flux ($\lambda < 4\text{--}5\ \mu\text{m}$), and a thermal component ($\lambda > 4\text{--}5\ \mu\text{m}$) corresponding to the absorbed solar flux, re-emitted at longer wavelengths (in addition, in the case of giant planets, the possible contribution of an internal flux). The solar contribution shows its maximum at about $0.5\ \mu\text{m}$, and maximum of the thermal contribution ranges from $7\ \mu\text{m}$ (in the case of Mercury) to $70\ \mu\text{m}$ (in the case of Neptune). Emission fluorescence can also be observed in the UV (especially in the Ly α line of hydrogen at $1216\ \text{\AA}$), the visible or the near-IR range (CH_4 in the giant planets at $3.3\ \mu\text{m}$, CO in Mars, Venus, Uranus and Titan at $4.7\ \mu\text{m}$).

In the case of the reflected sunlight component, planetary atmospheric signatures are observed in absorption in front of the solar spectrum. CO_2 (and

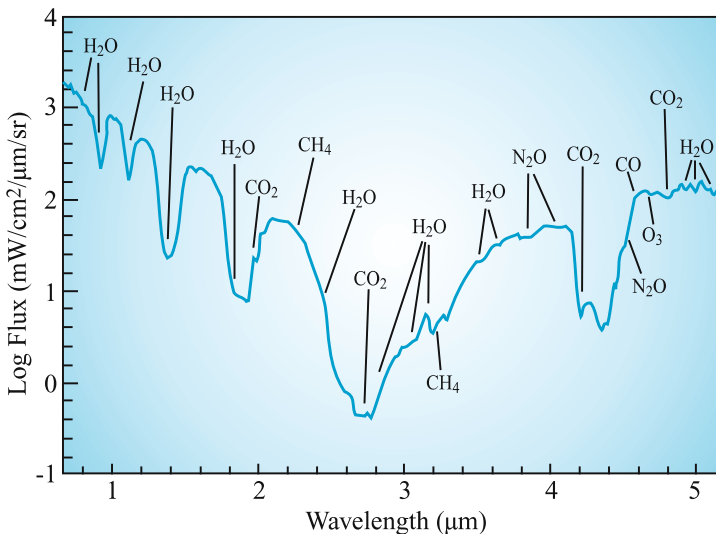


Fig. 5.7. The near-infrared spectrum of the terrestrial atmosphere, as recorded by the NIMS spectrometer on the Galileo spacecraft when it flew by the Earth in December 1990. Most of the absorptions are due to H_2O and CO_2 (after Drossart et al. 1993)

also H_2O in the case of the Earth) dominates the spectrum of terrestrial planets (Fig. 5.7) while CH_4 is predominant in the case of the giant planets (Fig. 5.8). Information is retrieved upon the column density of the atmospheric constituent, and shows, to first order, little dependence upon the atmospheric pressure and temperature. These measurements have been used to determine the mixing ra-

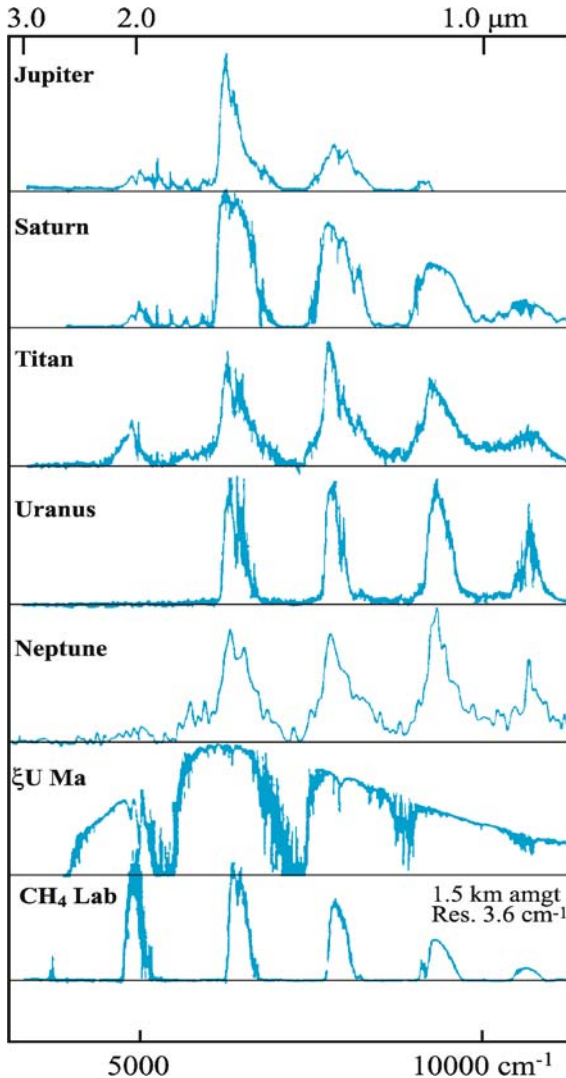


Fig. 5.8. Spectra of the giant planets in the near-infrared range, compared with a laboratory spectrum of CH_4 . A stellar spectrum is shown to indicate the absorptions due to the terrestrial atmosphere. It can be seen that the spectra of the giant planets are all dominated by methane (after Larson 1980)

tios of H_2O and CO in Mars, as well as NH_3 in Jupiter and CH_4 in the four giant planets. In the latter case, the penetration level, between the CH_4 absorption bands, can reach the troposphere down to about 1–3 bars. This method has allowed for the determination of the C/H ratio in all giant planets (Gautier and Owen 1986; Baines et al. 1995), as well as D/H in Mars from the study of the HDO band at $3.8\mu\text{m}$ (Owen et al. 1988; Krasnopolsky et al. 1997).

In the case of the thermal component, the outgoing flux strongly depends upon the thermal profile of the atmosphere. As mentioned above (Sect. 5.1.1) giant planets are characterized by a temperature inversion at the tropopause level (0.1 bar) where the temperature ranges from 110 K (Jupiter) down to 50 K (Uranus and Neptune). Molecular species seen below this level, in the troposphere, are observed as absorption features while stratospheric species, above the tropopause, are seen as emission features (Fig. 5.9; see Encrenaz 2000, for a review). No permanent inversion takes place in the middle atmospheres of Mars and Venus, so that atmospheric species, under nominal conditions, are observed in absorption in both the reflected and the thermal regime. In the case of Mars,

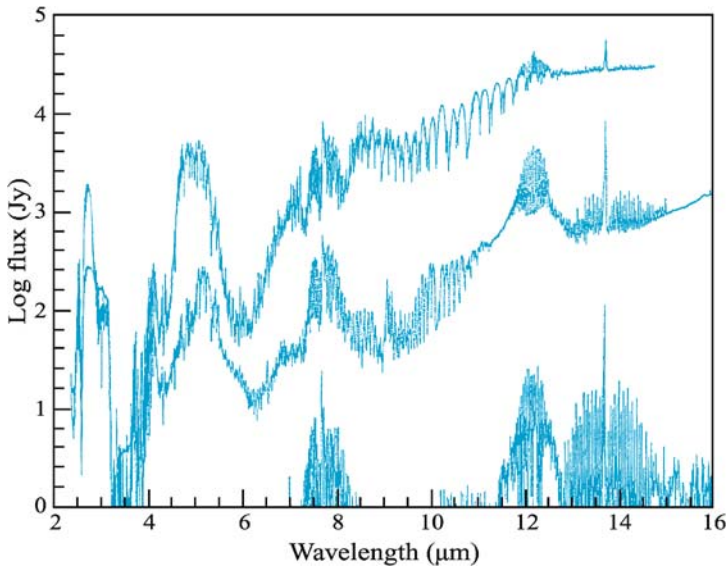


Fig. 5.9. The infrared spectra of Jupiter (*top*), Saturn (*middle*) and Neptune (*bottom*) observed by the SWS spectrometer of the ISO satellite. The reflected solar flux dominates below $4\mu\text{m}$, while thermal emission is observed at higher wavelengths. Molecular features appear there in emission, if formed in the stratospheres (CH_4 at $7.7\mu\text{m}$, C_2H_6 at $12\mu\text{m}$, C_2H_2 at $13.7\mu\text{m}$) or in absorption, if formed in the troposphere (NH_3 at $10\mu\text{m}$ in Jupiter, PH_3 at $8\text{--}9\mu\text{m}$ in Saturn). The flux of Uranus and Neptune is less than 1 Jy below $7\mu\text{m}$. Uranus (not shown in the figure) is detected only in the C_2H_2 emission band at $13.7\mu\text{m}$, with a flux equal to one third that of Neptune (after Encrenaz 1999)

however, a temperature inversion takes place over the polar caps, as the lower atmosphere is warmer than the surface; this inversion is observed in the profile of the strong CO_2 band at $15\mu\text{m}$ (Hanel et al. 1992; Fig. 5.10).

The thermal component has been used to identify most of the minor species of the giant planets, and to determine their mixing ratios (see Table 5.3). Data have come from ground-based and millimeter high-resolution spectroscopy, as well as from space data (IRIS/Mariner 9 on Mars, IRIS/Voyager on the giant planets and Titan).

In the case of Venus, in the absence of reflected sunlight, the thermal component can be detected down to very short wavelengths (about $1\mu\text{m}$) because the surface temperature is very high. This component has been observed on the night side of Venus, both from the ground and from space (Venus flyby of the Galileo spacecraft in February 1990). These data have been used to measure the abundances of minor species in the lower troposphere of Venus (CO , H_2O , HDO , HCl , SO_2 , OCS) and to determine the D/H ratio (Bézard et al. 1990; Fig. 5.11). The very high value (over 100 times the terrestrial value) inferred from these data, is strong evidence for an outgassing of large amounts of water early in the history of the planet. Indeed, the large present abundance of HDO , relative to H_2O , is interpreted as differential gravitational escape of the two molecules: HDO , being heavier, escapes at a slightly slower rate than does H_2O .

Over the past decade, a new generation of instruments has been developed, which couples both imaging and spectroscopic capabilities. These imaging spectrometers operate in the IR range ($1\text{--}5\mu\text{m}$ or $7\text{--}13\mu\text{m}$) on large (4-meter or 8-meter class) telescopes and offer a spectral resolving power ($\lambda/\delta\lambda$, λ being the

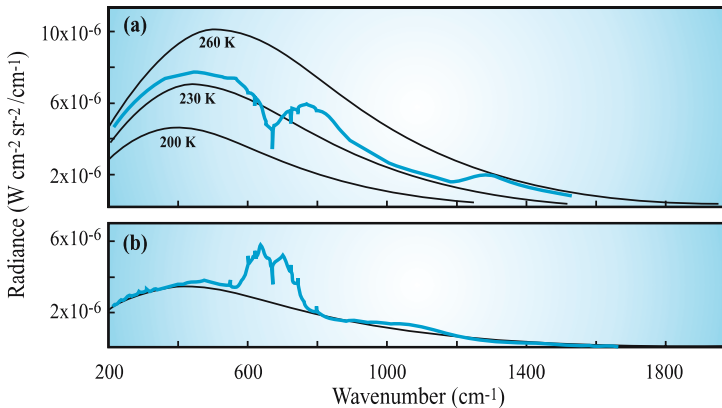


Fig. 5.10. The thermal spectrum of Mars obtained by the IRIS spectrometer onboard the Mariner 9 probe. (a) Middle latitudes. (b) South Pole. In the first case (a), the surface is warmer than the atmosphere, whereas the opposite applies above the pole (b). As a result, the CO_2 band centered at 660cm^{-1} is observed in absorption in (a), and in emission (with an inversion structure at the center reflecting the temperature profile) in (b) (after Hanel et al. 1992)

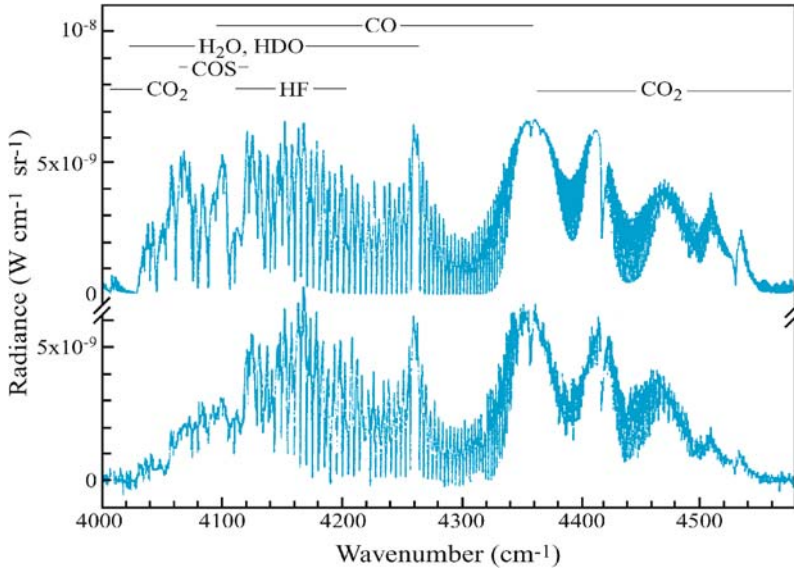


Fig. 5.11. The thermal spectrum of the night side of Venus, observed in the near-infrared range by high-resolution ground-based spectroscopy. *Top*: synthetic spectrum, calculated using an atmospheric model of Venus. *Bottom*: the observed spectrum. Molecular absorbers are indicated at the top of the diagram (after Bézard et al. 1990)

wavelength and $\delta\lambda$ the spectral resolution element) of 10^4 – 10^5 , with a typical spatial resolution of 0.3–0.5 arcsec, adapted to a good seeing. An even higher spatial resolution (down to the diffraction limit, i.e., 0.06 arcsec at $2.2\mu\text{m}$ for an 8-meter telescope) is now achievable with adaptive optics techniques, but is limited to bright objects if high spectral resolution is required.

In the millimeter/submillimeter range, it is possible to obtain a very high spectral resolving power (10^6), over a limited bandwidth (0.5 to 1 GHz), with the use of a specific technique called heterodyne spectroscopy. The spatial resolution is moderate in the case of a single antenna (10 arcsec at 230 GHz, or $\lambda = 1.3\text{mm}$, for the IRAM 30-meter antenna), but is much higher in the case of interferometers (IRAM, OVRO and, in the future, ALMA).

5.3.2 *In Situ* Analysis: Chemical Composition from Mass Spectrometry

In situ analysis has been used extensively for deriving the chemical composition of the neutral atmospheres of Mars and Venus. The method was fully successful on Mars with the mass spectrometry experiments aboard the Viking landers. Measurements of Venus (especially with the Venera landers) were more difficult, because of the extreme pressure and temperature conditions in the environment of the landers at the surface of Venus. Complementary with remote sensing

analysis, a coherent picture of the composition of the neutral atmospheres of terrestrial planets could be built (Table 5.3).

In the case of Jupiter, in situ mass spectrometry measurements have provided a major contribution about the neutral atmosphere of the Jovian atmosphere, thanks to the Galileo probe, which entered the planet on December 7, 2000.

Other in situ measurement techniques, based upon X and/or γ spectrometry, have been also very powerful for measuring elemental abundances at or under the surface of terrestrial planets. This has been true, in particular, in the case of the Moon and Mars. In 2002, γ -ray spectroscopy aboard the Mars Odyssey orbiter has shown evidence for large abundances of hydrogen atoms (presumably associated to water ice) under the surface of the Martian polar caps (Boynton et al. 2002; Bibring et al. 2004).

5.4 From Solar System Planets to Exoplanets

5.4.1 Properties of Detected Exoplanets: a Summary

Companions around stars have been searched for over the past decades. Since the discovery of many exoplanets around nearby stars, a classification has been set up between exoplanets ($>13 M_J$, i.e., 13 Jovian masses), brown dwarfs (between 13 and 74 M_J) and stars (beyond this mass). The limit between exoplanets and brown dwarfs corresponds to the ignition of the deuterium cycle of nuclear reactions, which lasts for about ten million years. Brown dwarfs were probably formed, like other stars, by gravitational collapse of a molecular cloud, but their temperature was not high enough to rise the full cycle of thermonuclear reactions that occur in stars.

In October 2005, about 170 exoplanets were detected (for a review see Chap. 6 by M. Ollivier), mostly by the velocimetry method. In a few cases, transits were also observed, leading to an independent determination of the planet's diameter and mass.

From the present data set of detected exoplanets, the following main characteristics can be derived:

- A large fraction of nearby stars have at least one large companion ($>1 M_J$ at 1 AU, $>2 M_J$ at 5 AU). About 5% of these stars have a companion of mass between 0.5 and 5 M_J at asteroisimologic distances less than 2.5 AU. We still have no information about “exoterrestrial planets” too small to be detected with the present means.
- There seems to be a correlation between the number of planet(s) around a star and its metallicity: the proportion of heavy elements in stars having planets seems to be higher than it is for stars without planets.
- The mass distribution of exoplanets shows a continuous distribution for objects larger than 10 M_J , and a number of objects increasing toward the lower masses. Still, the observational bias should favor the higher masses. We still

- do not know whether this trend extends down to terrestrial-type exoplanets. The present distribution seems to show a gap in the region 30–50 M_J (the so-called brown dwarfs desert).
- There is an accumulation of large exoplanets with periods of about 3 days ($R_s = 0.04\text{AU}$) and low eccentricities, with a quasiabsence of objects at shorter periods. Exoplanets with larger periods tend to have larger eccentricities.

5.4.2 Earth-Like Exoplanets: the Habitability Zone

Earth-like exoplanets (EEPs) cannot be detected yet, because ground-based techniques are not sensitive enough, in view of the low contrast between the exoplanet and its neighboring star. The COROT mission, to be launched by CNES in 2006, should provide us with the first detection of Earth-like exoplanets.

The variety of spectra exhibited by the terrestrial planets illustrates that EEPs might show very different spectra, depending upon their atmospheric composition and thermal structure, and the possible nature of their surface. Observing these spectra will obviously require very high imaging capabilities coupled with spectral devices, as currently under study in the frame of the DARWIN and TPF projects.

The presence of water (at $6\mu\text{m}$), methane (at $7\mu\text{m}$) and ozone (at $9.6\mu\text{m}$) in the mid-infrared spectrum of the Earth (Fig. 5.12) illustrates the interest of these spectral features as possible diagnostics for habitable EEPs and even for inhabited EEPs (L  ger 1997; Selsis 2002).

By analogy with the situation we know on Earth, in which the development of life was largely favored by the presence of liquid water, scientists have defined a region, the “habitability zone”, where the temperature is likely to allow the presence of liquid water and hence, possibly favor the development of life (see Chap. 7 by F. Selsis).

The presence of liquid water depends upon the surface temperature, which is only constrained, in the absence of the greenhouse effect, by the surface albedo and the planet’s distance from the star. As compared to the case we know in our solar system, where the habitability zone is close to 1AU from the Sun, the distance of the habitability zone to its star can greatly vary as a function of the star’s spectral type and mass. Hotter and massive stars will have very remote habitability zones, while in the case of small and cold stars; these regions will be located in the close vicinity of the stars.

It should be kept in mind, however, that the concept of the habitability zone might be misleading. In the case of our solar system, the habitability zone, as described above, would only include the Earth and possibly Mars. However we know that other sites are potentially interesting for emerging life: the putative underground ocean of Europa, one of Jupiter’s Galilean satellites, and the atmosphere of Titan, Saturn’s biggest satellite, where prebiotic molecules have been

found. The concept of a habitability zone is thus a useful tool for exobiology, but it could also appear to be too restrictive.

5.4.3 Giant Exoplanets: Structure and Composition

As mentioned above, most of the exogiant planets (EGPs) detected so far have been found in the close vicinity of their stars. Many of them have very small astero-centric distances (0.04–0.05 AU). In this case, the structure and composition of their atmospheres are likely to be strongly affected by the presence of

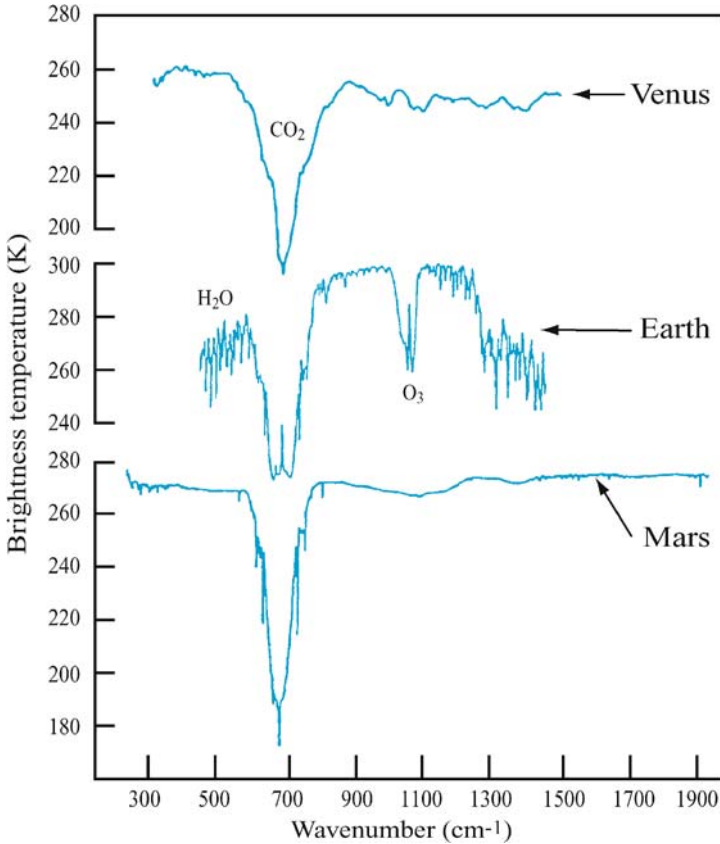


Fig. 5.12. The spectra of the terrestrial planets in the middle infrared (5–100 μm), obtained from Venera 15 (Venus), Nimbus 4 (Earth) and Mariner 9 (Mars). The CO₂ band at 15 μm dominates all spectra. The H₂O lines are strong on Earth and appear very weakly on Mars. The signature of ozone, centered at 1042 cm⁻¹ (9.6 μm) is clearly visible in the Earth spectrum. Assuming that life on exoplanets would lead, as on Earth, to the production of O₂, the O₃ spectroscopic signature (easier to detect than the one of O₂) could be a diagnostic for the search for life on extrasolar planets (after Hanel et al. 1992)

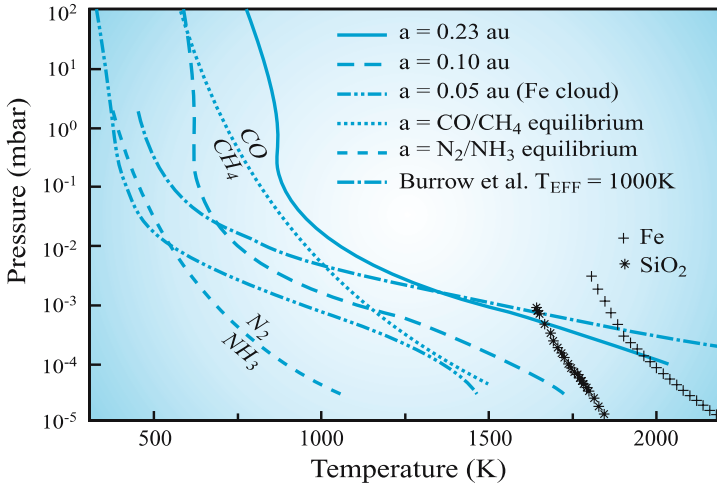


Fig. 5.13. The thermal structure of an EGP close to its star, for three different values of its astero-centric distance. The *curves* corresponding to thermal/chemical equilibrium for CO/CH₄ and N₂/NH₃ are also shown, as well as the condensation curves of iron and SiO₂. The parameter refers to the distance of the exoplanet to its star, expressed in astronomical units (AU) (after Goukenleuque et al. 1999)

the stellar wind, which, at 0.05 AU, is expected to be about 400 times stronger than the one arriving on Earth, for a solar-type star of one solar mass. This strong effect might even blow out the external atmospheric layers of the EGP (Vidal-Madjar et al. 2004).

The atmospheric structure of the EGPs has been studied by several authors, with and without the effect of the illuminating stars (Allard et al. 1997; Baraffe and Chabrier 1997; Guillot et al. 1997; Burrows et al. 1997; Goukenleuque et al. 2000). Under thermochemical equilibrium, and assuming cosmic abundances, the chemical composition of the atmospheres can be inferred. Besides hydrogen and helium, CH₄ and NH₃ are expected to be the major carbon and nitrogen-bearing molecules at stellar distances larger than 0.5 AU. N₂ is expected to prevail over NH₃ for stellar distances smaller than 0.2 AU, and CO should dominate over CH₄ for a stellar distance smaller than 0.06 AU (Goukenleuque et al. 2000). An important issue is the formation of condensates and the pressure level at which it takes place. Possible condensates are Mg₂SiO₄ at 2 bars, MgSiO₃ at 10 bars, and Fe at 10 bars (Goukenleuque et al. 1999; Fig. 5.13).

Many authors have calculated synthetic spectra of EGPs (Marley et al. 1996; Seager and Sasselov 1998; Sudarsky et al. 2000; Goukenleuque et al. 2000; Fig. 5.14). All results point out the interest of the 4–5 μ m and 10 μ m windows, which are relatively free of major atmospheric absorption, and where the contrast between the stellar and planetary fluxes is minimized. From the ground, the 4–5 μ m range, in particular, allows for the search of CO₂ (4.2 μ m) and CO

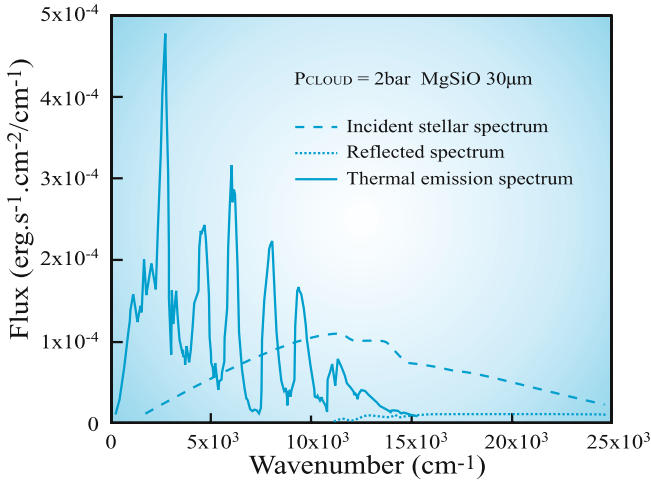


Fig. 5.14. The synthetic spectrum of an EGP of a type similar to 51 Peg b (after Goukenleuque et al. 1999)

($4.7\ \mu\text{m}$), while the $7\text{--}13\ \mu\text{m}$ range allows for the search of CH_4 ($7.7\ \mu\text{m}$) and NH_3 ($10.5\ \mu\text{m}$ space, the $6\text{--}10\ \mu\text{m}$ is best suited for the search for water ($6.2\ \mu\text{m}$) and O_3 ($9.5\ \mu\text{m}$).

5.5 Conclusions

This review has illustrated the very large variety of physicochemical conditions taking place in the atmospheres of the solar system planets. These conditions all reflect different processes, specific of each planet, of its formation and evolution.

With Venus and the Earth, we have seen that a relatively small difference in heliocentric distance can lead to drastically different evolution scenarios. The cases of Mercury and Mars illustrate the importance of the initial mass, both to provide significant internal energy (as a smaller mass leads to a smaller amount of radioactive elements in the core) and to capture the impacting bodies, which will contribute to the atmosphere. While Mercury and the Moon were too small and too warm to keep a stable atmosphere, Mars was indeed able to build an early dense atmosphere, but its internal energy vanished and the atmosphere rarefied.

The exploration of the giant planets has also shown unexpected differences between the two gaseous giants Jupiter and Saturn, and between the two icy giants, Uranus and Neptune, which are far from being understood. Whereas their global properties are well explained in the frame of the nucleation theory, the strong variations observed in their internal energies, and in their dynamical activities, remain a puzzle. The next step to better understand the diversity of

the giant planets would be a multiprobe mission, which would achieve in situ measurements of the four atmospheres: the Galileo probe has illustrated the huge potential of this exploration in the case of Jupiter, and the same type of information was obtained on Titan's atmosphere with the descent of the Cassini-Huygens probe in 2005.

Extrapolating the information from solar system planets to exoplanets is difficult, in view of the large diversity of atmospheric characteristics observed in the solar system. In addition, we have seen that the formation scenario of the solar system appears to be very different from the ones encountered by the exoplanets presently detected.

With the present techniques (velocimetry, transits) we are now in the exploration phase devoted to the detection of exoplanets and to the determination of their physical parameters (mass, density, radius). After the ground-based detection of many EGPs, we can hope, in the near future, to reach the detectability limit of EEPs with the space missions COROT, Kepler and possibly Eddington. Other high-angular resolution techniques, using photometry, imaging and spectroscopy, will have to be used to characterize the atmosphere of exoplanets. In all cases, the infrared range appears well suited for exploring the chemical composition and the atmospheric structure of exoplanets, due to the favorable contrast of the planet with respect to its star.

References

- Allard F., Haultschild P. H., Alexander D. R., Starrfield S. (1997). Model atmospheres of very low-mass stars and brown dwarfs. *Ann. Rev. Astron. Astrophys.* **35**, 137–177.
- Atreya S. K. (1986). *Atmospheres and Ionospheres of the Outer Planets and Their Satellites*. Springer, Berlin Heidelberg New York.
- Atreya S. K., Romani P. N. (1985). Photochemistry and clouds of Jupiter, Saturn and Uranus. In: *Recent Advances in Planetary Meteorology*, pp. 17–68. Cambridge University Press, Cambridge.
- Atreya S. K., Wong M. H., Owen T. C. et al. (1999). A comparison of the atmospheres of Jupiter and Saturn: deep atmospheric composition, cloud structure, vertical mixing, and origin. *Planet. Space Sci.* **47**, 1243–1262.
- Atreya S. K., Mahaffy P. R., Niemann H. B., Wong M. H., Owen, T. (2002). Composition and origin of the atmosphere of Jupiter: an update, and implications for the extrasolar planets. *Planet. Space Sci.* **51**, 105–112.
- Baines K. H., Mickelson M. E., Larson L. E., Ferguson D. W. (1995). The abundances of methane and ortho/para hydrogen on Uranus and Neptune: implications of New Laboratory 4-0 H₂ quadrupole line parameters. *Icarus* **114**, 328–340.
- Baraffe I., Chabrier G. (1997). Theory of very low-mass stars, brown dwarfs and extrasolar giant planets. In: Paresce F. (ed.) *Science with the VLT Interferometer*, pp. 72–79. ESO/Springer, Berlin.
- Bibring J.-P. et al. (2004). Perennial water ice identified in the south polar cap of Mars. *Nature* **428**, 627–630.
- Bergstrahl J. T., Miner E. D., Matthews M. S. (eds.) (1991). *Uranus*. University of Arizona Press, Tucson.

- Bézard B., de Bergh C., Crisp D., Maillard J.-P. (1990). The deep atmosphere of Venus revealed by high-resolution nightside spectra. *Nature* **345**, 508–511.
- Boynton W. V., Feldman W. C., Squyres S. W. et al. (2002). Distribution of hydrogen in the near-surface of Mars: evidence for subsurface ice deposits. *Science* **297**, 81–85.
- Burrows A., Marley M. S., Hubbard H. B. et al. (1997). A non-gray theory of extrasolar giant planets and brown dwarfs. *Astrophys. J.* **491**, 856.
- Coustenis A., Taylor F. W. (1999). *Titan, the Earth-Like Moon*. World Scientific, New York.
- Coustenis A., Lorenz R. (1999). Titan. In: Weissman P.R. et al. (eds.) *Encyclopedia of the Solar System*, pp. 405–434. Academic, New York.
- Cruikshank D. (1995). *Neptune and Triton*. University of Arizona Press, Tucson.
- Davies J. (2001). *Beyond Pluto*. Cambridge University Press, Cambridge.
- de Pater I., Romani P. R., Atreya S. K. (1991). Possible microwave absorption by H₂S in Uranus and Neptune’s atmospheres. *Icarus* **91**, 220–233.
- Drossart P. et al. (1993). Earth global mosaic observations with NIMS-Galileo, *Planet. Space Sci.* **41**, 551–561.
- Encrenaz T. (1999). The planet Jupiter. *Astron. Astrophys. Rev.* **9**, 171–219.
- Encrenaz T. (2000). ISO observations of solar system objects. In: Casoli F. et al. (eds.) *Infrared Space Astronomy, Today and Tomorrow*, pp. 89–149. EDP Sciences/Springer, Berlin.
- Encrenaz T. (2002). Atmospheric structure, composition and diagnostics. In: *Earth-Like Planets and Moons. ESA SP 514*, 183–190.
- Encrenaz T., Drossart P., Feuchtgruber H. et al. (1999). The atmospheric composition and structure of Jupiter and Saturn from ISO observations: a preliminary review. *Plan. Space Sci.* **47**, 1225–1242.
- Encrenaz T., Bibring J.-P., Blanc M., Barucci M.-A., Roques F., Zarka P. (2004). *The Solar System*. Springer, Berlin Heidelberg New York.
- Fegley B., Gautier D., Owen T., Prinn R. G. (1991). Spectroscopy and chemistry of the atmosphere of Uranus. In: Bergstrahl J.T. et al. (eds.) *Uranus*, pp. 147–203. University of Arizona Press, Tucson.
- Feuchtgruber H., Lellouch E., de Graauw T., Bézard B., Encrenaz T., Griffin M. (1997). External supply of oxygen to the atmospheres of the giant planets. *Nature* **389**, 159–162.
- Gautier D., Owen T. (1989). The composition of outer planet atmospheres. In: Atreya S.K. et al. (eds) *Origin and evolution of planetary and satellite atmospheres*, pp. 487–512. University of Arizona Press, Tucson.
- Goukenleuque C., Bézard B., Lellouch E. (2000). A radiative equilibrium model of 51 Peg-type planets. In: Griffith C., Marley M. (eds.) *From Cool Stars to Planets*, p. 242, ASP Conference Series, San Francisco.
- Goukenleuque C., Bézard B., Joguet B., Lellouch E., Freedman R. (2000). A radiative equilibrium model of 51 Peg b. *Icarus* **143**, 308–323.
- Graedel T. E., Crutzen P. J. (1993). *Atmospheric Change, an Earth-System Perspective*. Freeman, New York.
- Guillot T., Marley S. M., Saumon D., Freedman R. S. (1997). In: Eiroa C. et al. (eds.) *Infrared Space Interferometry: Astrophysics and the study of Earth-like planets*, pp. 37–46. Kluwer, Dordrecht.
- Hanel R. A., Conrath B. J., Jennings D. E., Samuelson R. E. (1992). *Exploration of the Solar System by Infrared Remote Sensing*. Cambridge University Press, Cambridge.

- Hunten D. et al. (1984). Titan. In: Gehrels T. et al. (eds.) *Saturn*, pp. 671–759. University of Arizona, Tucson.
- Krasnopolsky V. A. (1986). *Photochemistry of the Atmospheres of Mars and Venus*. Springer, Berlin Heidelberg New York.
- Krasnopolsky V. A., Bjoraker G. L., Mumma M. J., Jennings D. E. (1997). High-resolution spectroscopy of Mars at 3.7 and 8 μm : a sensitive search of H_2O_2 , H_2CO , HCl , and CH_4 , and detection of HDO. *J. Geophys. Res.* **102**, 6525–6534.
- Larson H. P. (1980). Infrared spectroscopic observations of the outer planets, their satellites, and the asteroids. *Ann. Rev. Astron. Astrophys.* **18**, 43–75.
- Léger A. (1997). Life signatures on exoplanets. In: Eiroa C. et al. (eds.) *Infrared Space Interferometry: Astrophysics and the Study of Earth-Like Planets*, pp. 47–54. Kluwer, Dordrecht.
- Lellouch E., Belton M., de Pater I., Paubert G., Gulkis S, Encrenaz T. (1992). The structure, stability and global distribution of Io's atmosphere. *Icarus* **98**, 271–295.
- Lunine J. I. (1999). *Earth: Evolution of a Habitable World*. Cambridge University Press, Cambridge.
- Marley M. S., Saumon D., Guillot D. et al. (1996). Atmospheric, evolutionary, and spectral models of the brown dwarf Gliese 229B. *Science* **272**, 1919–1921.
- Marten A., Gautier D., Owen T. et al. (1993). First observation of CO and HCN on Neptune and Uranus at millimeter wavelengths and their implications for atmospheric chemistry. *Astrophys. J.* **406**, 285–297.
- Mizuno H. (1980). Formation of the giant planets. *Progress of Theoretical Physics* **64**, 544–557.
- Noll K. S., Weaver H. A., Feldman P. D. (1996). *The Collision of Comet Shoemaker-Levy 9 and Jupiter*. Cambridge University Press, Cambridge.
- Owen T., Encrenaz T. (2003). Element abundances and isotope ratios in the giant planets and Titan. In: Kallenbach R. et al. (eds.) *Solar System History From Isotopic Signature of Volatile Elements*. *Space Science Reviews* **106**, 121–138.
- Owen T., Mahaffy P., Niemann H. B. et al. (1999). A low-temperature origin for the planetesimals that formed Jupiter. *Nature*, **402**, 269–270.
- Pollack J. B., Hubickyj O., Bodenheimer P., Lissauer J., Podolak M., Greenzweig Y. (1996). Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124**, 62–85.
- Ruzmaikina T. V. (1997). In: Shirley J.S., Fairbridge R.W. (eds.) *Encyclopedia of Planetary Sciences*. Chapman and Hall, London.
- Rosenqvist J., Lellouch E., Romani P., Paubert G., Encrenaz T. (1992). Millimeter-wave observations of Saturn, Uranus and Neptune: CO and HCN on Neptune. *Astrophys. J.* **392** L99–L102.
- Seager S., Sasselov D. D. (1998). Extrasolar giant planets under strong stellar irradiation. *Astrophys. J. Lett.* **502**, 157–160.
- Selsis F. (2002). Search for signatures of life on exoplanets. *ESA SP* **514**, 251–258.
- Stern S. A., Tholen D. J. (eds.) (1997). *Pluto and Charon*. University of Arizona Press, Tucson.
- Spitzer, L. (1952). The terrestrial atmosphere above 300 km. In: Kuiper G. (ed.) *The Atmospheres of the Earth and Planets*, pp. 211–247. University of Chicago Press, Chicago.
- Sudarsky D., Burrows A., Pinto P. (2000). Albedo and reflection spectra of extrasolar giant planets. *Astrophys. J.* **538**, 885–903.

- Vidal-Madjar A., Désert J.-M., Lecavelier des Etangs A. et al. (2004). Detection of oxygen and carbon in the hydrodynamically escaping atmosphere of the extrasolar planet HD 209458 b. *Astrophys. J.* **604**, L609–L612.
- West R. A. (1999). Atmospheres of the giant planets. In: Weissman, P.R. et al. (eds.) *Encyclopedia of the Solar System*, pp. 315–337. Academic, New York.

6 What About Exoplanets?

Marc Ollivier

This text aims at synthesizing what the discovery of the first extrasolar planets – 147 in January 2005¹ – has brought to modern planetology and exoplanetology. In particular, we will present an overview and will consider the numerous questions that are raised by present observations. We will not detail the methods for detection of these planets but will only summarize them (Sect. 6.4 and Appendix 1.1). Several papers have already been devoted to that subject.² Finally, taking into account the interdisciplinary public to which this paper is devoted, a chapter containing an astrophysical overview and additional information has been added in Appendix 1.1. To get a more complete picture of modern astrophysics and to understand its vocabulary, see Léna 1996 (in French) or Maran 1992.

6.1 Let's Talk About History

The question of the existence of other planets, whether they are inhabited or not, in the Universe, is certainly one of the oldest questions raised by astronomers and arose as soon as the analogy between our Sun and other stars was made. We can find traces of this debate in antiquity. To quote Epicurus, mainly known for his atomistic theories, who claimed in his “Address to Herodotus”: “It is not only the number of atoms, it is also the number of worlds which is infinite in the Universe. There is an infinite number of worlds, similar to ours and an infinite number of different worlds [...] One must agree that in all these worlds, without any exception, there are animals, plants and all the living beings we observe.”³ Even if this text should be read keeping in mind the antique vision of what Epicurus calls “the world”, these words remain avant-gardism at an epoch where the whole universe was described by the criteria of aestheticism proposed by geometry.

¹ The list of exoplanets is reproduced in Appendix 1.3.

² An easy reading synthesis for beginners can be found in the first chapter of Ollivier's dissertation (in French, Ollivier 1999) and a reference and figure list can be found in the exoplanet encyclopedia (Schneider 2004).

³ This quotation, like others in this chapter, is translated from Mayor and Frei's book published in 2001. This books (in French) offers an excellent historical introduction to this subject.

For several thousand years, this debate stayed in the domain of personal feeling, and as a consequence, of dogmatic fight when the personal feeling contradicted the official dogma, and particularly, the dogma of the powerful Church in the Middle Ages. When he published in 1585 “The infinite, the Universe and the Worlds,” Giordano Bruno, “the agitator”, claimed in public some ideas which began to rise up counter to the official dogma. He paid the price for such words and was condemned in 1600 to the stake by the Inquisition. A few years later, Huygens (1629–1695) reformulated very clearly these ideas and wrote: “Do not hesitate, us, to admit like other main philosophers of our epoch that the nature of stars and the sun is the same. From that point results much grandiose conception of the world than the conception, which corresponds to past views more or less traditional. Since what prevents now from thinking that each of these star or Suns has planets around and that these planets are endowed with moons?”

We must wait until the beginning of the twentieth century, to see progress in the field of astronomical instrumentation that would allow a new scientific and observational approach to this question, and hence started the race towards the discovery of objects with ever lower masses. Techniques similar to those used for the discovery of the first planets brought the first observational elements to the discussion: the Doppler effect, radial velocimetry and astrometry. However, immaturity of the methods and a poor control of experimental biases did not allow a quick confirmation of the first announcements, rather leaving the scientific field with a strange feeling of science fiction. In 1931, Berman announced the discovery of a 45 Jovian mass object (M_{Jup}). In 1963 Van de Kamp announced the detection, by astrometry (observation of the central star movement) of a 1.6 Jupiter mass planet, located at 4.42 astronomical units (AU)⁴, on a high eccentricity orbit ($e = 0.77$) with a period of about 24 years around Barnard’s star (a star located at 5.9 light-years⁵ from our Sun). Two years later, a new analysis of his data let him to conclude that two objects were orbiting Barnard’s star. This discovery, which has not yet been confirmed by radial velocity measurements, is however the illustration of the efforts expended at that epoch to reach observational results. Van de Kamp compared the position of Barnard’s star on 2413 pictures taken with the 80cm Sproul Observatory telescope between 1916 and 1963, i.e., on a period of 50 years where everything has more or less changed starting from photographic plates to the instrumental drifts.

The advent of high angular resolution coupled with high dynamic techniques (speckle interferometry, adaptive optics and coronagraphy) at the beginning of the 1980s brought a batch of observations but no real discoveries. Beta Pictoris has been observed since 1981. It exhibits strong infrared excess, revealing the presence of a dusty disk; one can then asks if it could be a planetary system

⁴ The astronomical unit (AU) is equal to the mean distance between the Sun and the Earth. It is about 150 million kilometers. This unit is adapted to the description of planetary systems.

⁵ 1 light-year = distance the light ($v \sim 300\,000\text{ km s}^{-1}$) travels in one year = 9.46×10^{12} km.

in formation. In 1981 the Geneva Observatory recorded a sudden drop in the Beta Pic light curve: was it the signature of a planetary transit or the evidence of the existence of comets? The Beta Pic case is strong. In 1985 McCarthy and his collaborators announced the discovery of a giant planet around the star Van Biesbroeck 8 using speckle interferometry techniques. The announcement caused much fuss. The refutation based on new observations by Perrier and Mariotti in 1987 was somewhat hidden. In 1992, Wolszczan and Frail announced the discovery of three planets including one whose size was comparable to our Moon, around the pulsar PSR 1257 +12 (Wolszczan and Frail 1992). The discovery passed a bit unnoticed because the announcement was made exactly the same day as another team published a refutation of a similar discovery around the pulsar PSR 1829-10. These planets around pulsars were not of great interest because the electromagnetic environment would leave little chance for the hypothetical development of life on them. Planet researchers got confused once more when, in 1995, Walker and his team announced that they had not detected any planets larger than $3 M_{\text{Jup}}$ (their instrumental limit) around 21 stars in the solar neighborhood.

The continuation of the story is well-known. The efforts towards instrumentation development, and particularly concerning the radial velocimetry technique, lead in 1995 to the announcement of the discovery of the first extrasolar planet, 51 Peg b, around a solar-like star. This discovery was done by Mayor and Queloz (Mayor and Queloz 1995).

6.2 Statistical Analysis of the First Extrasolar Planets Discoveries

By the end of January 2005, 147 giant extrasolar planets (Wolszczan's three planets not taken into account in this list), grouped into 128 planetary systems including 15 multiple systems (for instance, three giant planets may orbit around the star Upsilon in the Andromeda constellation) have been detected, mainly by the radial velocity method (cf. Appendix 1.2) already used to detect 51 Peg b (Schneider 2004). The initial list of these objects can be found in Appendix 1.3. The detection of these planets is the result of the observation of about 4000 stars in the solar neighborhood (a few tens of parsecs (pc)⁶). The observed stars are mainly dwarfs belonging to the main sequence⁷ and of spectral type F, G, K, M ($3000\text{K} < T_{\text{eff}} < 7500\text{K}$). Massive stars (spectral type O, B) are removed from

⁶ The parsec (pc) is a unit of distance. It corresponds to the distance from which the Sun-Earth system (separated by 1 AU, about 150 million kilometers) is observed at an angular separation of 1 arc second. $1 \text{ pc} = 3.085678 \times 10^{16} \text{ m} = 3.261633 \text{ light-years}$. The parsec is particularly suited to the question of extrasolar planets, because an analogous Earth around a star at N pc will be seen under an angle of $1/N$ arc second.

⁷ See Appendix 1.1 for a definition of the spectral classification.

the sample because hot stars exhibit less spectral features than colder stars. In addition, as they rotate quickly the accuracy of the radial velocimetry method is limited. Finally, the short lifetime of these stars (several hundred million years at best) does not allow forming of evolved planetary systems and these stars are thus less interesting from the exobiology point of view. The time sampling and the observational history vary a lot among the sample of 4000 stars. Some stars have been observed from the beginning of the radial velocity measurement programs (1994), others have been recently added when new instruments have allowed an increase in the number of potential targets. It also should be mentioned that the 4000 stars have been observed by several teams all over the world, each working on its own sample, and that the redundancy between these samples allowed the confirmation or refutation of candidates by other teams. Finally, we estimate that the present giant planet discovery announcement rate is about a dozen objects per year.

We will now interest ourselves in these planets and will try to synthesize what their discovery teaches us about their formation, their evolution and planetary system dynamics. To meet this goal, we will favor a statistical approach based on the study of mass, distance, orbit eccentricity and metallicity of the parent star distributions. Then, we will focus on several particularly well-observed objects, for instance, HD 209458 b, which is a giant planet that exhibits a transit in front of its parent star, thus allowing for a more detailed study.

6.2.1 The Mass Distribution of Exoplanets

The planet detection method, by radial velocity measurement (cf. Appendix 1.1), is not specific to the planet search but also allows the detection of stellar companions. When one measures $M \cdot \sin(i)$ (i is the angle at which the system is observed, with $i = 90^\circ$ if the system is seen edge-on), one gets a lower limit to the companion mass. Before we focus more specifically on exoplanets, it is interesting to look at the companion mass distribution, whatever the companion may be, star or planet (Fig. 6.1).

It is clear that the distribution is bimodal, with planets ($M < 0.01 M_\odot$ ⁸) and stars ($M > 0.08 M_\odot$). Between these two regions, there are very few objects. This zone was identified in the 1980s and is known as the brown dwarfs desert. It was detected in the first surveys dedicated to the search for substellar companions, as applied to a few tens of objects, and it was concluded that there was a lack of this kind of object (Campbell et al. 1988; Marcy and Benitz 1989). Present radial velocity surveys allow the identification of only about 20 brown dwarf candidates (for which the measured values of $M \cdot \sin(i)$ are included in the brown dwarf mass range). However, the use of the European astrometric Hipparcos satellite has eliminated, with great confidence, seven objects out of the 11 candidates detected

⁸ $1 M_\odot =$ mass of the Sun

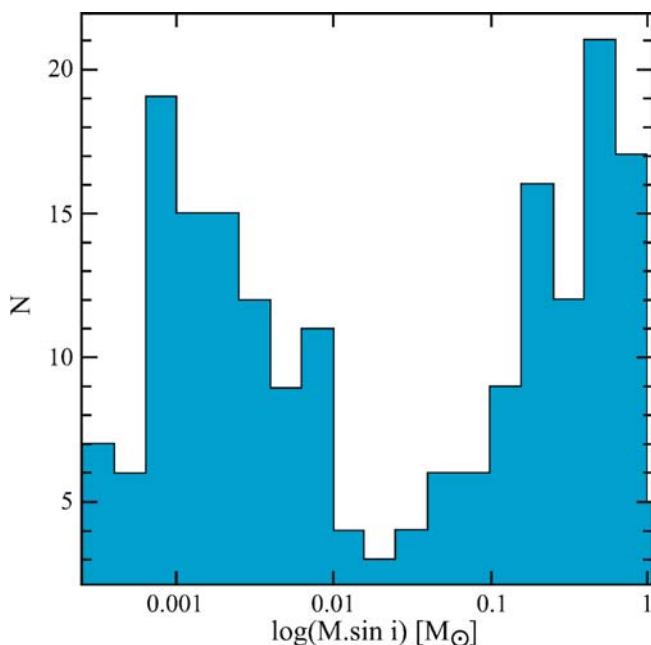


Fig. 6.1. Histogram of companions $M \cdot \sin(i)$ in the case of binary systems. One clearly sees that the distribution is bimodal with, *on the left*, the planets, and *on the right*, the stars. The *middle zone* ($0.01 M_{\odot} < M \cdot \sin(i) < 0.08 M_{\odot}$) is known as the brown dwarfs desert (from Santos et al. 2002)

by the ELODIE instrument⁹ (Halbwachs et al. 2000). This last result tends to show that not only is the brown dwarfs desert a reality, but that also brown dwarf candidates detected by the radial velocimetry technique are generally objects with a greater mass, for which $\sin(i)$ is small. This important separation between planet and star can be explained, or at least understood, if we consider standard formation models. Stars form by gravitational collapse of a more massive cloud (few $100 M_{\odot}$) while planets, even giant planets, form by material accretion within a planetary disk. In the first case, one goes from massive to less massive by fragmentation, in the second case, one goes from planetesimals to planets by accumulation and accretion. The two processes have different control parameters that can explain different limit scales.

Let us examine in more detail the mass distribution of exoplanets. Figure 6.2a and Fig. 6.2b represent, respectively, the histogram of $M \cdot \sin(i)$ and the masses derived by statistical deconvolution for known planets.

The mass distribution of exoplanets peaks towards low mass objects ($M < 5M_{\text{Jup}}$). It is not an observational bias: the angle deconvoluted figure (Fig. 6.2b)

⁹ Astrometric data from Hipparcos would allow a solution to the determination on $\sin(i)$ values.

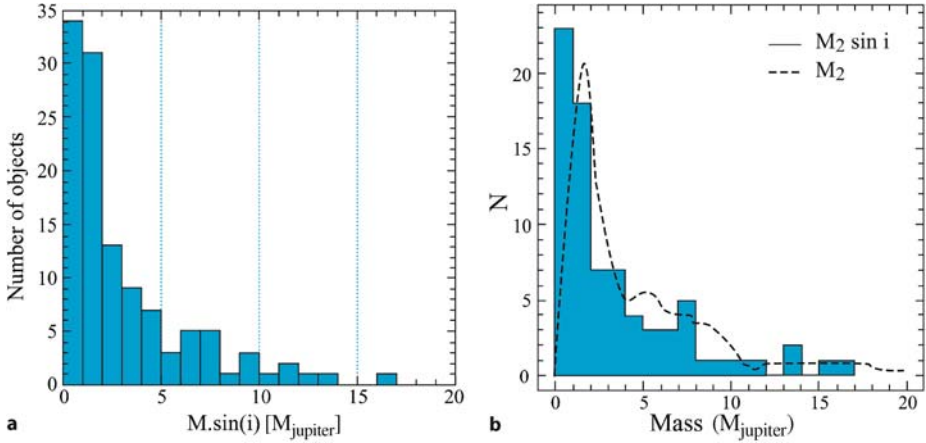


Fig. 6.2. **a** Histogram of $M \cdot \sin(i)$ for 119 of 147 presently known exoplanets. **b** Histogram for 88 planets (*in blue*) deconvoluted by the $\sin(i)$ distribution assuming the equiprobability of i values (after Santos et al. 2002)

shows the same result. On the other hand, and as shown in Appendix 1, the radial velocity measurement detection method is very sensitive to massive objects at short distances from their parent stars. If more massive objects had existed, the method should have detected them, even at large distances from their stars (several AU). This means that the giant planet formation process seems a priori to allow for, or at least to make easier, the formation of relatively low mass objects with short periods. One also notices that the more one looks for long period objects, the more massive the detected objects are (Udry et al. 2003). Better statistics are necessary to establish strong hypotheses.

6.2.2 The Star Planet Distance Distribution

Figure 6.3 shows the distance distribution histogram of the presently known exoplanets from their parent stars. Figure 6.3b clearly shows a bimodal distribution for the exoplanet distances, with the following observations:

- Planets with very short distance from their parent stars, typically 0.05 AU: the *hot Jupiters*¹⁰
- A significant lack of objects at about 0.3 AU
- Planets at a distance larger than 0.3 AU, uniformly distributed

Once again, it should be mentioned that the lack of planets at 0.3AU is not a bias of the method, which is very sensitive to objects near their parent stars. The measurement of the distance to the star is performed by the orbital period measurement (period of the radial velocity variation) and the application

¹⁰ Planets with the size of Jupiter, near (0.05 AU) their parent star.

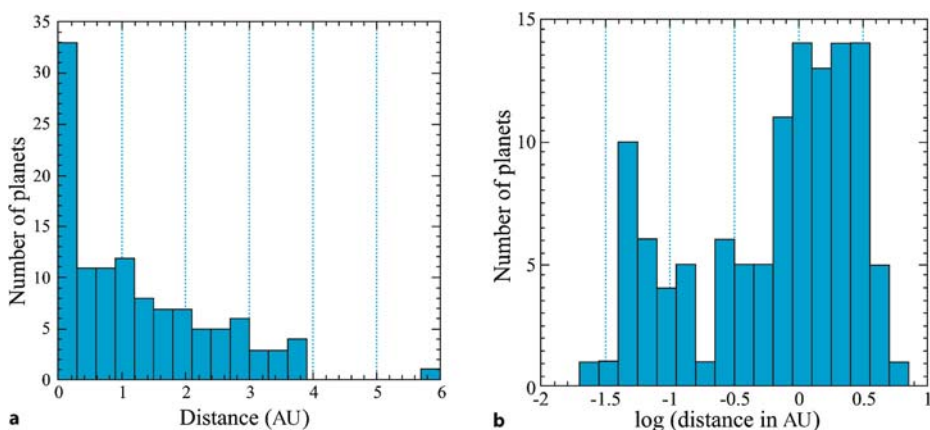


Fig. 6.3. **a** Distance distribution of 119 of 147 known exoplanets. **b** Same data but with a logarithmic distance scale

of Kepler’s third law (cf. Appendix 1.1). The spectral analysis gives a good knowledge of the central star properties and, particularly, of its mass.

This bimodal distance distribution shows that the objects are not uniformly distributed and that some stable orbits are favored. Some of them are very close to the star. One can search for some explanations of this observation in the planet formation process (see Part 7 in Black and Matthews 1985). One can also consider the possibility that a planetary orbit varies during the planet’s lifetime. One must then consider the hypothesis of orbit migrations. We will now discuss this point.

6.2.3 Orbit Migration

Even if we can find in the literature authors who propose potential mechanisms to explain planetary formation at very short distances (Wuchterl 2000), then the evocation of orbit migration mechanisms is necessary to explain the hot Jupiters. Several points should be mentioned here:

- The condensation of refractory materials (for instance, silicates that are components of cores) is only possible at temperatures below 1500K. However, present planetary models predict that the protoplanetary disk temperature at 0.05 AU (present position of hot Jupiters) is about 2000K. As a conclusion, hot Jupiters cannot form at their present position. In classical giant planet formation models, the formation limit for a Jupiter-like planet around a solar-like star is about 5 AU, which is the ice condensation limit under the primitive disk conditions. The temperature of the disk at that point is about 160K.
- With the same protoplanetary disks models, the dust density is too low to form a solid giant planet core at 0.05 AU.

- The gas density at 0.05 AU is also too low to allow for the formation of a giant planet starting from an existing core.
- Planetary evolution models predict that the giant planets have a diameter 10 times larger when they form than at a later stage of their formation. If this point is true, this also means that the gas escape velocity on such a planet is reduced when the planet forms. In other words, the planet can easily and quickly evaporate before it reaches its final state.

If hot Jupiters do not form at their present position, and migrate, one must find one or several mechanisms allowing for this phenomenon. Several hypotheses have been proposed.

6.2.3.1 Interactions Between Two Giant Planets

This mechanism was proposed by Rasio and Ford (1996). It supposes two giant planets in the system, initially at a large distance from their central star. To allow for interactions the planets must satisfy the Hill's stability criterion, which can be written as the following, for circular orbits:

$$\frac{|a_1 - a_2|}{a_1} < 2.4 \left(\frac{M_1}{M_*} + \frac{M_2}{M_*} \right)^{1/3} \quad (6.1)$$

where a_1 , a_2 , M_1 , M_2 are the semi-major axes of the orbits of planets 1 and 2, their corresponding masses, and M_* is the mass of the parent star. Under such conditions, the direct collision of the two planets has a probability of about 50%. In this case, a single object is formed (mass $\leq M_1 + M_2$) without any change of the distance from the star, but on a high eccentricity orbit. If no collision occurs, a set of successive interactions between the planets can lead to an angular momentum transfer from one to another. The studies, calculations and numerical simulations by Rasio and Ford show two main cases:

- One of the two planets is ejected from the system and the gravitationally attached remaining planet migrates towards the star and gets stabilized on an orbit with a period larger than about 0.4 year. This orbit has a strong eccentricity. This model could be invoked for HD 80606, which has an eccentricity of 0.927, but its period near 112 days is in contradiction with the minimum value of the orbital period (0.4 year) required by this mechanism.
- If the distance between one of the planets and the star becomes less than about 0.05 AU, the orbit of this planet can be circularized by gravitational interactions with the star. In this case, the other planet is stabilized on a more external orbit without being ejected. This model thus predicts that all the hot Jupiters must belong to multiple systems with a giant planet at a large distance. This last point is a good test for the model, that future astrometric observations (sensitive to massive objects at large distances) should confirm or negate in the next few years. It should be mentioned that the future

astrometry missions GAIA (ESA) and SIM (NASA) should allow for, with microarcsecond accuracy measurements (at least for SIM), the identification of objects to the limit of 1 Earth mass at the distance of Jupiter.

6.2.3.2 Interaction with a Distant Binary Companion

This model has mainly been developed by Eggleton and Kiseleva-Eggleton (2001). It brings into play two successive phenomena:

- The so-called Kozai’s interaction mechanism (Kozai 1962) that allows an angular momentum transfer between a planet and a distant stellar companion. This mechanism supposes that the initial orbit of the planet is inclined with respect to the stellar companion orbital plane. The result of this interaction is a high eccentricity orbit for the planet. Eggleton and Kiseleva-Eggleton show that, starting from an initial inclination angle of 39 degrees, the eccentricity of the planet orbit reaches 0.51. If this angle is maximum (90 degrees), then the resulting eccentricity is near 1.
- A mechanism of orbit circularization by tide effects. Because the planet is transferred to a high eccentricity orbit, the periastron is in the neighborhood of the star and tidal effects are important. The characteristic duration of the orbit circularization has been estimated by the authors to be 10^7 years for a final period of 2.1 days.

From a theoretical point of view, this migration process is not one of the simplest. However, it can be invoked in certain cases, and particularly when the binary companion has been identified.

6.2.3.3 Angular Momentum Transfer Between a Planet and a Disk

This mechanism is by far the most studied one. It is also the one for which the models have been extended to explain the positions of the planets that have been detected. This mechanism was proposed in 1979, when no exoplanet had been yet discovered (Goldreich and Tremaine 1979). It is not necessary to present here all the details and subtleties of this model, we’ll just give the general principles and essential consequences. Many papers go into the subject more closely (Lin et al. 1996; Ward 1997; Trilling et al. 1998, 2002; Nelson et al. 2000; Kuchner and Lecar 2002; Armitage et al. 2002). This mechanism requires that the planet forms quickly in the disk and interacts with it before the disk is dispersed during the T Tauri phase of the star. It thus concerns a priori only fast forming planets, i.e., giant planets.

The “Hill radius” around a planet is the limit upon which the gravitational attraction of this planet becomes lower than the attraction of the parent star. The Hill radius, written R_H is given by the following relation:

$$R_H = \left(\frac{M_p}{3M_*} \right)^{1/3} a \quad (6.2)$$

where M_p is the mass of the planet, M_* is the mass of the star and a is the semi-major axis of the planet orbit. The angular momentum transfer between the disk, which is supposed to be Keplerian¹¹, and the planet allows for migration through the effects of a localized interaction inside the Hill radius around the planet. The planet experiences a torque due to a double contribution:

- The gas inside the planetary orbit orbits faster than the planet itself and thus gives angular momentum to the planet before it falls toward the star and accelerates. The gas is therefore at the origin of a force that pushes the planet outward its orbit.
- The gas outside the planetary orbit orbits slower than the planet itself, and takes angular momentum from the planet and escapes from the planetary orbit neighborhood. This gas is at the origin of a torque that opposes the torque produced by the inner gas and drags the planet toward the star.

Because of Lindblad¹² resonances proximity, the two torques are not equal. The torque from the external gas is stronger and the net effect is a loss of angular momentum. The planet thus migrates toward the star. Two types of mechanisms can be distinguished:

1. *Type I migrations:*

They concern low-mass planets for which one gets typically:

$$\frac{M_p}{M_*} < 3 \left(\frac{H}{a} \right)^3 \approx 0.1 M_{\text{Jup}} \quad (6.3)$$

where H is the disk thickness. In that case, one gets $R_H < H$ and the planet does not create a gap in the disk. The engine for the mechanism is the differential Lindblad torque. Its characteristic time τ_1 (in years) can be written (Terquem et al. 2000) as

$$\tau_1 = 10^{10} \left(\frac{M_p}{M_{\text{Earth}}} \right)^{-1} \left(\frac{\sigma}{\text{g} \cdot \text{cm}^{-2}} \right)^{-1} \left(\frac{a}{\text{AU}} \right)^{-1/2} \left(\frac{H}{a} \right)^2 \quad (6.4)$$

where σ is the disk density, H its thickness, M_p the mass of the planet and a its semi-major axis. It is interesting to note that in this case the migration speed of the planet linearly increases with respect to the mass and the density of the disk. However, the migration speed does not depend upon the viscosity of the disk. One point is not taken into account in this model: the turbulence in the disk. Nelson et al. (2000) show that the turbulence

¹¹ The dynamics of this disk is governed by Kepler's laws. Particularly, the angular speed distribution is not uniform in the disk but depends on the distance from the star.

¹² Position where a resonance phenomenon between the density waves in a disk and the motion of planets in this disk happens. The consequence is an energy transfer from the waves to the planets.

has effects that can quickly become comparable in terms of energy to the proposed mechanism of type I migrations. It is thus very difficult to ignore this phenomenon in the description. Taking into account the low limit value of the mass of the mechanism ($\sim 0.1 M_{\text{Jup}}$), we can not expect immediate observational experiments that will confirm the reality of this mechanism.

2. *Type II migrations:*

They concern more massive planets following the relation

$$\frac{M_p}{M_*} > 3 \left(\frac{H}{a} \right)^3 . \quad (6.5)$$

In this case, the Hill radius of the planet is bigger than the disk thickness, so that the planet creates a gap in the disk, thus cleaning a zone in the disk with a width about twice that of the Hill radius one of the planet. This mechanism is illustrated in Figs. 6.4 and 6.5. Two opposing effects occur: gravity tends to increase the gap (tide effects), and disk viscosity tends to close the gap and homogenize the disk.

The consequence of this competition is that the migration time of the planet depends only on the disk viscosity and not on its mass or on the mass of the planet. The migration time τ_{II} (in years) for a type II mechanism is given by the following expression (Terquem et al. 2000):

$$\tau_{\text{II}} = 0.05 \frac{1}{\alpha} \left(\frac{a}{H} \right)^2 \left(\frac{a}{\text{u.a.}} \right)^{3/2} . \quad (6.6)$$

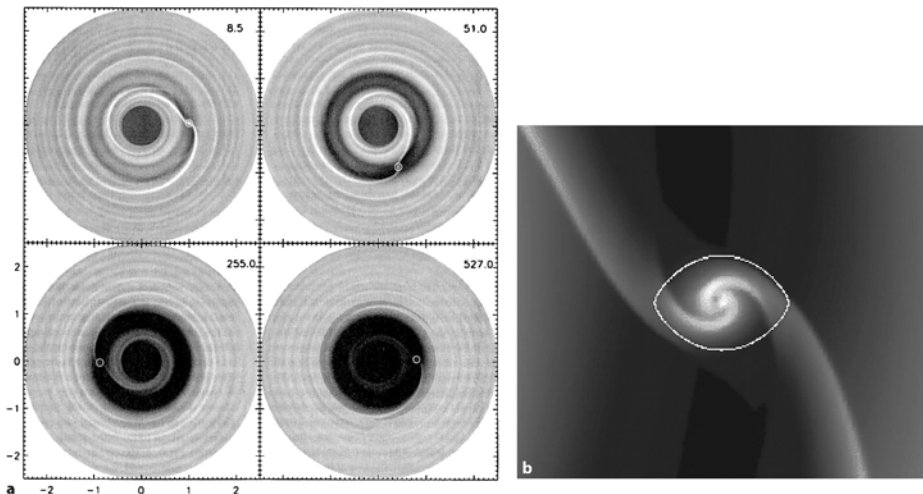


Fig. 6.4. **a** Formation of a gap during the migration of a giant planet in a disk, according to a type II mechanism (Nelson et al. 2000). **b** Direct neighborhood of the planet

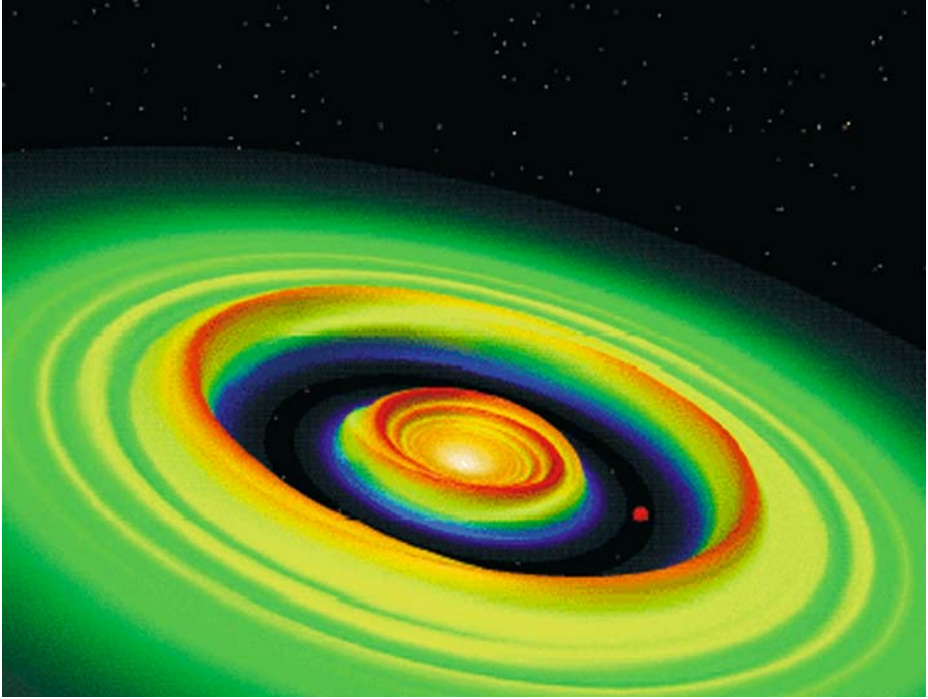


Fig. 6.5. Numerical simulation of the structure of a disk during the migration of a giant planet according to a type II mechanism (Bryden and Lin 1999)

We note that a type I migration (without the opening of a gap) is faster than a type II migration (with the opening of a gap). Figure 6.6 allows for a comparison of the migration time corresponding to each mechanism as a function of the planetary mass.

The main drawback of this mechanism of migration by angular momentum transfer between a planet and a disk is the lack of a mechanism to stop the process. However, several aspects need to be considered:

- The short distance tide effects lead to a transfer of angular momentum from the star to the planet. This angular momentum stops the planets (mechanism comparable to the mechanism that drags the moon away from the Earth).
- There are effects of the magnetic field at the inner edge of the disk.
- The mass losses at the Roche¹³ lobe (Trilling et al. 1998). At a distance of a few stellar radii from the star (star Roche lobe) mass can escape from the planet and fall onto the star. The mass loss reduces the migration

¹³ The Roche lobe of an object is the zone in which the gravitational attraction of the object itself is higher than the gravitational attraction of the central star. All the material inside the Roche lobe is stable in the planetary neighborhood.

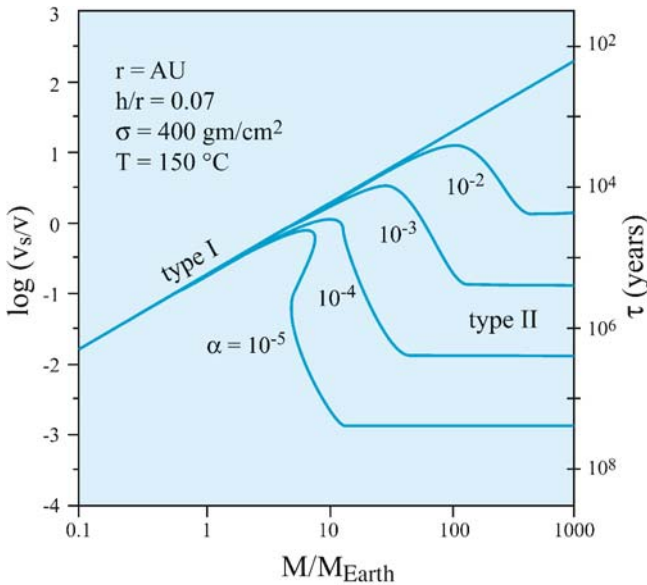


Fig. 6.6. Numerical simulation of the structure of a disk during the migration of a giant planet according to a type II mechanism (Bryden and Lin 1999)

mechanism efficiency. This phenomenon leads to a distinction between planets with masses lower than M_{Jup} , that migrate quickly towards their star and lose all their mass by escape at the Roche lobe, and more massive planets that can survive this loss of mass, which stops when the disk dissipates (typical time 10^7 years). The planet then evolves in the stellar neighborhood.

- The important stellar wind for pre-main sequence stars. In order for this mechanism to be efficient, one should consider a hypothetical component parallel to the disk that would be oriented against the planetary movement and would stop the migration.
- The dissipation of the disk before the planet is absorbed by the star. This would require a perfect correspondence between the disk lifetime and the planetary migration time. This mechanism seems (a priori) completely ad hoc.
- The low density of the disk in the neighborhood of the star makes the migration mechanism inefficient.

6.2.3.4 Resonant Interactions with a Disk of Planetesimals

This mechanism has been proposed and studied by Murray et al. in 1998. It consists of an angular momentum transfer between a planet and planetesimals in which the trajectory (and particularly the eccentricity of the orbit) is modified.

This modification of planetesimal eccentricity is at the origin of their collision with the star, with the planet itself, or their ejection toward the infinity, and out of the solar system. The critical parameter of this model is the planetesimal density in the disk that directly governs the migration time and the final value of the semi-major axis of the migrating planet orbit. Figure 6.7 shows the efficiency of the migration as a function of the surface density of the disk. The surface density of planetesimals is given by: $\Sigma(r) = \Sigma_0(1\text{AU}/r)^{3/2}$ with several values of Σ_0 .

As shown in Fig. 6.7, the main limitation of this model is the high planetesimal density required to explain an efficient migration such as for a hot Jupiter. This mechanism does not explain, as it had been previously considered, the distance distribution of the observed objects.

6.2.3.5 Dynamic Friction with a Disk of Planetesimals

This mechanism is similar in principle to the preceding one. The friction between a giant planet and a disk of planetesimals leads to a centripetal force on the planet. The particularity of this mechanism is that it leads to circular orbits with semi-major axis bigger than 0.1 AU. Beyond this limit there are no longer enough planetesimals to maintain the mechanism. Once again this mechanism can be evoked marginally but is not sufficient to explain, by itself, the distance distribution of the observed objects. One can find the details of this model and numerical simulations of the orbits evolution as a function of time and mass of the disk in Popolo et al. (2001).

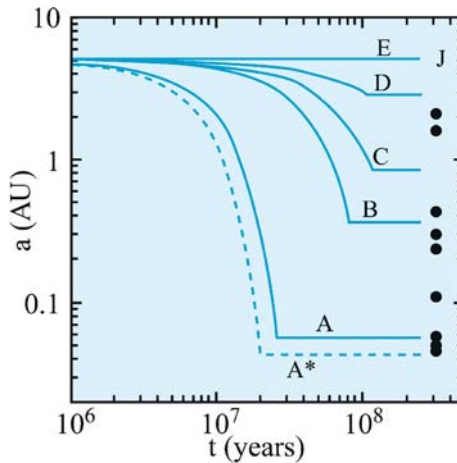


Fig. 6.7. Efficiency of the migration as a function of the surface density $\Sigma(r)$ for $\Sigma_0 = 8000$ (A), 2000 (B), 1200 (C), 600 (D) and 40 (E) g cm^{-2} and for a planet initially located at 5.2 AU on an orbit with an eccentricity of 0.048 (from Murray et al. 1998)

6.2.4 Mass/Distance Relation for Exoplanets

Figure 6.8 shows the distance of the detected exoplanets as a function of their mass. This diagram shows a correlation between the masses and distances of planets (Udry et al. 2003; Zucker and Mazeh 2002). Particularly:

- A sharp limit in the mass/distance distribution with an empty zone (hatched zone on Fig. 6.8)
- The lack of high mass objects at short distances
- The lack of low-mass objects ($M < M_{\text{Jup}}$) at distances between 0.5 and 5 AU

Taking into account the small number of objects included in this diagram, one immediately turns to an observational or a statistical bias to explain these points. An observational bias can quickly be eliminated because the hatched zone is precisely a zone where objects are easily detectable, taking into account the present sensitivity of the radial velocity detection methods. Udry et al. showed that the statistical bias is also very unlikely. Using a Monte Carlo simulation they showed that, considering the number of detected planets uniformly spread over the detectability zone, the probability that the hatched zone is effectively empty is only 0.00028. Their conclusion is thus that the lack of objects in this part of the diagram is the result of a real physical phenomenon.

From Sect. 6.2.3 on orbit migrations, one can correlate the mass of the object to the migration efficiency (even if in the case of type II migrations, the migration time is a priori only governed by the disk viscosity). Massive objects have larger inertia and in this case, the migration process is less efficient (Armitage et al. 2002). One can find a complete discussion of the mass effect on migration efficiency in Udry et al. 2003. Finally, Mazeh and Zucker pointed out a possible

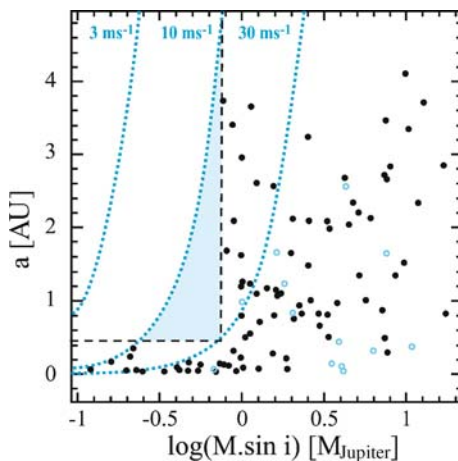


Fig. 6.8. Distance of the detected exoplanets as a function of their mass (Udry et al. 2003)

correlation between the masses and the orbital periods for the case of multiple planetary systems (Mazeh and Zucker 2003).

6.2.5 Eccentricity of Exoplanet Orbits

Figure 6.9a shows the eccentricity of the presently known exoplanets as a function of their orbital period (and thus more or less as a function of their distance from their parent stars). Figure 6.9b shows the orbits of all presently detected short period planets.

From these graphs two points can be raised:

1. For orbital periods shorter than 7 days the exoplanets orbits are roughly circular.
2. For orbital periods longer than 7 days, and opposite to what is observed in our solar system (solar system planets have low eccentricity orbits) there does not appear to be any correlation between eccentricity and orbital period. The distribution is comparable to that of stars in multiple systems.

Short period orbit circularization can be explained by tidal effects. If a planet is in the neighborhood of its parent star, the gravitational potential gradient creates a differential attraction between the inner side (oriented towards the star) and the outer side of the planet. The result is the formation of a tire-shaped structure at the planet surface. This tire-shaped structure is at the origin of the energy dissipation by internal friction within the planet that tends to circularize the orbit. The torque on this tire-shaped structure also tends to synchronize the rotation and revolution of the planet. The time

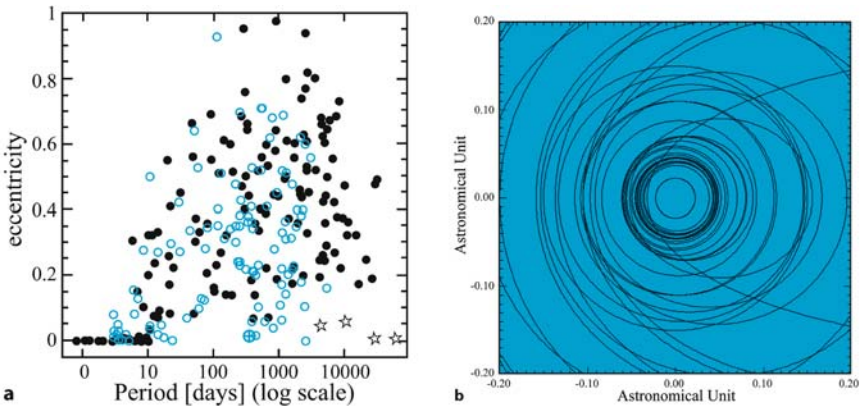


Fig. 6.9. **a** Eccentricity of exoplanets (*left*) (*black dots*) as a function of the distance from their parent star. For comparison, the *blue polygons* are stars in multiple systems, and the *star symbols* are the planets of our solar system (from Santos et al. 2002). **b** Orbits of short period planets (*right*) plotted on the same scale. Quasicircular orbits are observed for very short periods

required to circularize and to synchronize rotation and revolution is about one billion years for planets with orbital periods shorter than 7 days (Halbwachs et al. 2005). A consequence of this rotation/revolution locking is that the planet always shows the same side to the star (as in the case for the Earth/Moon system, the same side of the moon is always seen from the Earth). In the case of planets with an atmosphere, this irradiation asymmetry can lead to powerful zonal winds between the bright side (hot) and the dark side (cold). This last point depends strongly on whether the atmosphere is thermalized or not. There are still not enough observations to support this discussion.

As mentioned before, planets with periods longer than 7 days seem to have an eccentricity distribution comparable to that of stars in multiple systems. How can we justify a different mechanism for the two classes of objects? We can consider the fact that, at greater distances from the parent star, the interactions with the protoplanetary disk (Goldreich and Sari 2003) or with other planets and stars in the immediate neighborhood (Marzari and Weidenschilling 2002) becomes dominant compared to the tidal effects that tend to circularize the orbits. These interactions increase the system's eccentricity compared to the initial systems. As far as the solar system is concerned, the lack of massive objects in the solar neighborhood can explain the low eccentricity of the planetary orbits.

We must also mention that several theoretical models predict that the eccentricity of orbits decreases with the mass of the planet. Figure 6.10 shows that this hypothesis is not confirmed by the observations.

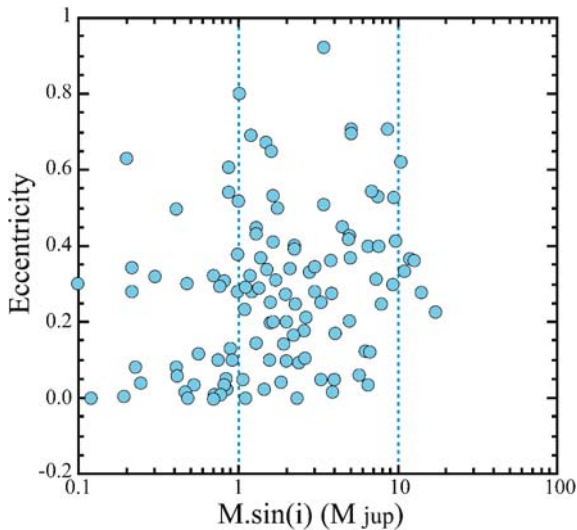


Fig. 6.10. Eccentricity of planetary orbits as a function of their mass. This diagram shows no evident correlation

6.2.6 The Metallicity of Stars with Planets

The periodic table of the elements can be summarized, by astronomers into three categories:

1. Hydrogen, whose relative mass abundance is written as X .
2. Helium, whose relative mass abundance is written as Y .
3. All heavier atoms are generically called metals, and their relative mass abundance is written as Z or Fe.

By definition, $X + Y + Z = 1$ and one calls Z the metallicity of a star. For instance, for the Sun, $X = 0.73$, $Y = 0.25$ and $Z = 0.02$ (these abundances are called the cosmic abundances). We can also define the metallicity (with no risk of confusion with Z), as

$$[Fe/H] = \log \left(\frac{[Fe][H]_{\text{sol}}}{[H][Fe]_{\text{sol}}} \right). \quad (6.7)$$

This value is equal to zero for stars with a solar metallicity, is negative (or positive) for stars less (or more) metallic than the Sun. We also say that stars with a metallicity higher than the Sun's (cosmic metallicity) exhibit an excess of metallicity.

Figure 6.11 is a histogram of the metallicities of stars with exoplanets. It clearly shows that there are more planets around stars with high metallicity. This result has been much discussed in the astronomical community. The difficulty in such a study is to correct the results for possible biases at each given level, particularly starting with the selection of the observation sample. This is the case for Fig. 6.11, which shows results that are

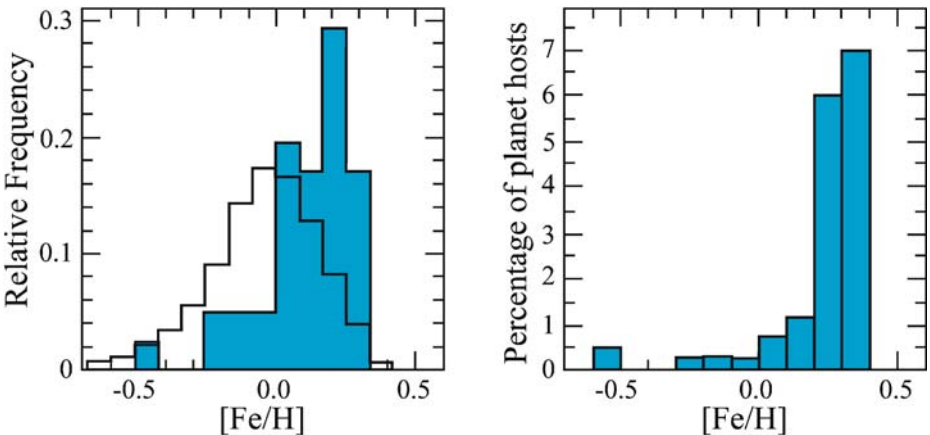


Fig. 6.11. *Left:* metallicity histogram for stars with planets (*blue*) compared to the rest of the sample (*white*). *Right:* relative fraction of stars with planets as a function of metallicity (after Santos et al. 2003)

now generally accepted. This result tends to show that giant planets form mainly around metal rich stars and are compatible with the classical model for planet formation by accretion of planetesimals and the formation of embryos from which the giant planets nuclei form. The question is now to understand the consequences of giant planet formation for telluric planets. The discovery of the first telluric planets by COROT, then by Kepler should allow a similar study for all categories of objects and should allow for a detailed study of the role of star metallicity in the process of planetary system formation.

6.3 The Atmospheres and Spectra of Giant Exoplanets

Up to now, no complete spectrum of an extrasolar planet has been observationally obtained. However, several chemical compounds have been identified in the atmosphere of HD 209458 b (see Sect. 6.3.2.2). In this section, we will discuss the proposed atmospheric models.

6.3.1 General Considerations

As for stars, the spectrum of giant exoplanets echoes the nature and the composition of the external atmosphere of the planet. However, and contrary to stars (at least hot stars) several condensed compounds such as water ice (solid H_2O), solid iron, forsterite, titanium calcium oxides (CaTiO_3) or titanium magnesium oxide (MgTiO_3) can contribute to the opacity of the atmosphere. The main difference between the models can be found in the thermalization rate of the atmosphere between the irradiated side and the dark side of the planet, and whether the role of dust and aerosol clouds are taken into account or not. We will focus on this point in Sect. 6.3.2.

Whatever the model may be, the spectrum of an exoplanet has two components:

Table 6.1. A spectral classification of exoplanets (Sudarsky et al. 2003)

Class	Distance	T_{eq}	Main chemical compounds	Remarks
1	a few AU	< 150 K	CH_4 , NH_3	IR flux low
2	1–2 AU	≤ 250 K	H_2O	H_2O bands
3	1 AU	350–800 K	H_2O , CH_4 , Na, K	Low albedo, no clouds
4	0.1 AU	1000 K	CO, Na, K, Li, Ru, H_2O	Silicates not visible
5	0.05 AU	≥ 1400 K	H_2O , CO, clouds	See Sect. 6.3.2

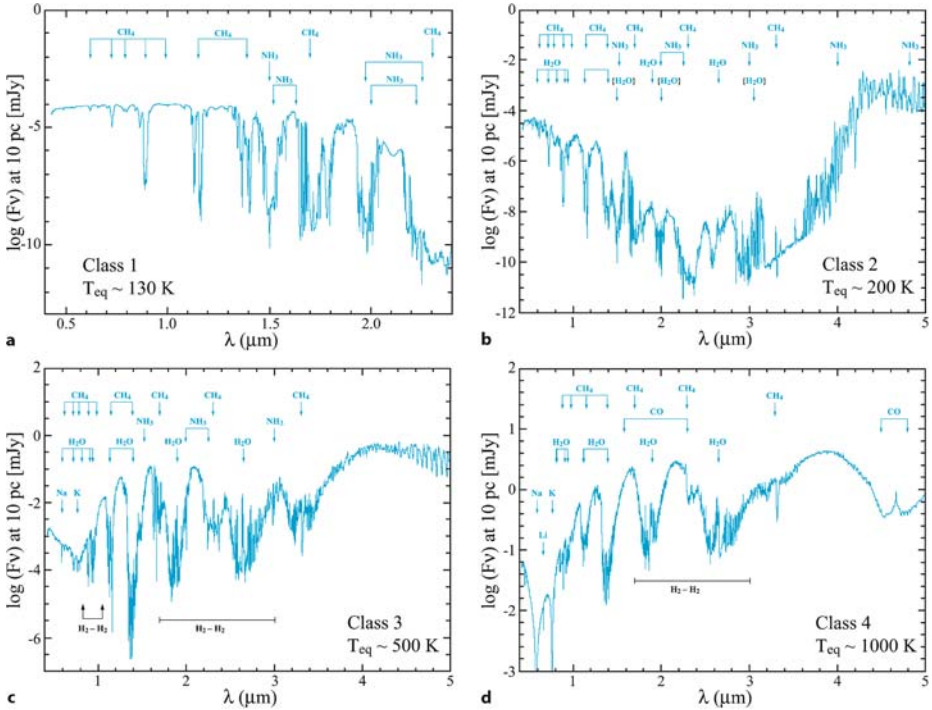


Fig. 6.12. Synthetic spectra of class 1, 2, 3, 4 giant exoplanets (Sudarsky et al. 2003)

1. A *reflective component* (in the visible and near-infrared spectral range) reproduces the spectrum of the star multiplied by the albedo (albedo = light reflection coefficient as a function of the wavelength).
2. A *proper emission component* (in the thermal infrared spectral range) depends on the temperature of the planet.

This temperature dependence led several authors to propose a theoretical spectral classification of planets as a function of their equilibrium temperature and thus to the distance from their parent star (Sudarsky et al. 2003). The main elements of this classification are shown in Table 6.1.

In the framework of this model, where spectral signatures are easily visible (not the case in the model of Barman et al. 2001), the role of clouds and aerosols is minimized. The spectra corresponding to classes 1, 2, 3, and 4 are shown in Fig. 6.12. The spectrum of a hot Jupiter (class 5) is shown in Fig. 6.13.

6.3.2 Pegasides: the Point of View of Theoreticians and Observers

6.3.2.1 Theoretical Models

The first theoretical models of Pegaside (the name given to hot Jupiters by analogy with the first object discovered in this category, 51 Pegasi b) inter-

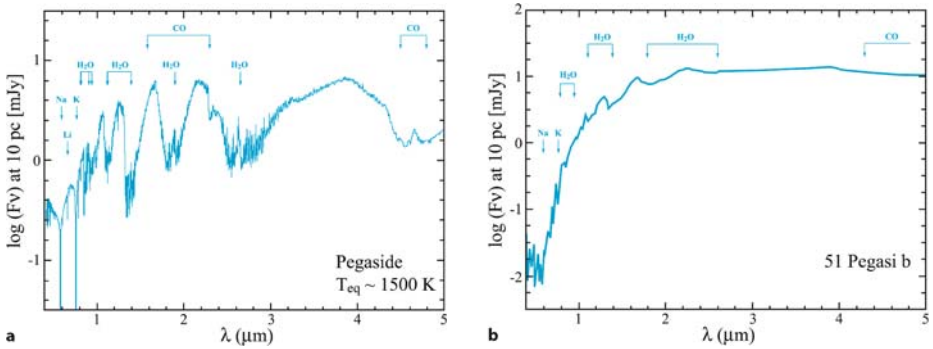


Fig. 6.13. Examples of synthetic spectra of a hot Jupiter (51 Peg b). **a** From Sudarsky et al. (2003). **b** From Barman et al. (2001), with same units and scale (planet at 10 pc). For comparison, the Sun at the distance of 51 Peg would have a flux of about 8 Jy at $3\mu\text{m}$

nal structure and atmosphere were developed just after the 51 Peg b discovery. They were aimed at an understanding of how such an object could be stable at such a short distance from the parent star. The first models were based on an effective temperature of 1250 K and showed that the Jean’s evaporation of the planetary atmosphere had no consequence for the planet’s survival (Guillot et al. 1996). Particularly, there was no mass transfer between the planet and the star, which means that the atmospheric radius was smaller than the Roche lobe of the planet. As mentioned in Sect. 6.2.5, because of their short period (short distance from the parent star) Pegasides have circularized and synchronized orbits (period of revolution = period of rotation), allowing several authors to predict the existence of very powerful zonal winds (typical speed greater than 1 km s^{-1}) between the bright and dark sides of the atmosphere (Showman and Guillot 2002). As a consequence, these zonal winds would redistribute the energy between the two hemispheres and partially thermalize the atmosphere.

The spectra of class 5 objects (in Sudarsky’s classification) predicted a strong component in the visible and near-IR spectral range corresponding to the light reflected from the nearby parent star and an infrared component associated with a black body at an effective temperature between 150 and 1500 K (depending on the model), with spectral features of H_2O and CO , but also Na and K . As mentioned before, the role of silicates and clouds of particles (aerosols) is still under discussion. Taking these clouds into account is fundamental for determining the thermalization rate of the atmosphere and plays a crucial role in an understanding of the spectral feature depths, associated with atoms and molecules. Several models have been proposed to describe the atmosphere of Pegasides (Sudarsky et al. 2003; Barman et al. 2001; Goukenleuke et al. 2000). Figure 6.13 shows two such model results.

Theoretical models for the formation of these objects assumes a fully convective structure for the planet in which the evolution has two phases (Guillot et al. 1996; Guillot and Showman 2002):

1. An initial formation with a fast contraction and an increase of the effective temperature.
2. A Cooling down, which is slower than for a classical giant planet (Jupiter) because of the irradiation (proximity of the parent star, at 0.05AU, i.e., 5% of the Sun–Earth distance). The consequence of this slow cooling down is the presence of a low thermal gradient (strong difference with Jupiter), which leads to an external radiative zone that slows the gravitational contraction.

The main conclusion of these models is that the radius of a Pegaside, of the same mass, should be larger than the radius of a weakly-irradiated planet such as Jupiter.

6.3.2.2 The Case of HD 209458 b

HD 209458 b is a hot Jupiter that was detected in 1999 around the star HD 209458 (Mazeh et al. 2000), a G0V spectral type main sequence dwarf, with a visible apparent magnitude of 7.65 in the visible spectral range and located about 47pc from the Sun. The period of the planet is 3.53474 days, which puts it at about 0.045AU from its parent star. Like all the other hot Jupiters, the probability for the planet to pass in front of its star (transit), assuming a random orientation for the planetary system compared to the observer on the Earth, is about 10% (see Appendix 1.2). HD 209458 b is one of the rare nearby objects (and the first that has been detected) that effectively transits in front of its parent star.¹⁴ The transit of HD 209458 b has been intensively observed from the ground and from space with the HST. Ground-based observations, supplemented by space measurements, have allowed a determination of the radius of the object in the visible spectral range assuming a broad band. This radius is equal to $1.347 \pm 0.06 R_{\text{Jup}}$ (Henry et al. 2000; Charbonneau et al. 2000; Brown et al. 2001) for a mass equal to $0.69 \pm 0.05 M_{\text{Jup}}$ (determined by radial velocity observations; in the case of a transit, the uncertainty in the value of $\sin(i)$ can be resolved). This observation is interesting because it does not confirm the theoretical models, mentioned in the previous paragraph, which predict a diameter smaller than that observed. These models however fit pretty well the parameters of the recently observed OGLE transit candidates, and particularly two new candidates (OGLE-TR-113 and OGLE-TR-132) that were recently published (Bouchy et al. 2004).

The spectroscopic study of the star HD 209458 before and during a transit of its companion allowed detecting, by a difference measure, several compounds in the atmosphere of the planet. This technique, which requires a good signal to

¹⁴ Several other candidates (5 up to now) including, for instance OGLE-TR 56, have been identified. Usually there are objects around fainter stars than HD 209458 and thus more difficult to study.

noise ratio and a good spectral resolution, can observe compounds either abundant in the atmosphere or with a large absorption cross section. The observation of the sodium doublet variations (around 589nm), using the STIS instrument on the HST, with a spectral resolution of more than 5000, showed that sodium is present at a concentration lower than expected (Charbonneau et al. 2002). This observation can be explained if one considers that the limb size is reduced by the presence of high altitude clouds. If this hypothesis is confirmed it has consequences both for the atmosphere, which is out of thermal equilibrium, and also for the full spectrum of the object. In addition, monochromatic observations showed that the radius of the planet, determined by the transit depth, depends on the wavelength. The opacity thus varies with the wavelength. One can find a detailed discussion of the radius of HD 209458 b and its atmosphere in Allard et al. (2003).

Observation in Lyman α (121.5nm), with the HST/STIS instrument showed that the band is attenuated by 15% during a transit (Vidal-Madjar et al. 2003a). At such a distance from the star (equal to about 8.5 times the stellar radius), the Roche lobe radius of HD 209458 b is about $3.6 R_{\text{Jup}}$. Vidal-Madjar et al. showed that, if the Roche lobe were fully filled with H_2 , the attenuation of the Lyman α band should not be more than 10%. Their conclusion is thus that HD209458b evaporates and loses hydrogen. This result contradicts the first model of hot Jupiters that excluded, with an effective temperature of about 1250K, any thermodynamic escape of the atmosphere. This problem can be solved if we consider that the temperature of the exosphere (outer zone of the atmosphere where the escape happens) is quite different from the effective temperature of the planet (Lammer et al. 2003; Vidal-Madjar et al. 2003b). Processes that can occur at this altitude are out of equilibrium processes. Exosphere temperatures of several thousands of K can thus been considered. Such temperatures can thus explain the slow evaporation of HD 209458 b. The evaporation rate has been estimated to be a few 10^{10} g s^{-1} (Lecavelier des Etangs et al. 2004). The lifetime of such an object would thus be about 10^{10} to 10^{11} years. In other words, after 5Myrs the planet would have lost between 1 to 7% of its mass by atmospheric escape. One question still remains. Are there objects for which this escape rate is much higher and that would have lost their entire atmosphere? In this case only the solid core of the giant planet would remain. One also must mention that the escape rate of hot Jupiters depends on the temperature of the central star. HD 209458 is a G0V spectral type star ($T \sim 6000\text{K}$). The result could be very different if we consider a colder star (K and M spectral type). Systematic observations of transits by future space missions (see Sect. 6.4) should bring new elements to this discussion.

These results should be reconsidered after the new observations by Vidal-Madjar et al. (2004) who, with the HST/STIS instrument, showed that not only hydrogen escapes but also oxygen and carbon. This point supposes that the whole atmosphere can evaporate, not only the light elements.

6.4 Future Steps in Exoplanetology and Associated Instrumentation

In reading the preceding pages, the reader will understand how it is difficult to conclude a paper on such a brand new topic. However, it is in this case necessary to summarize the relevant problems and open questions.

6.4.1 Open Questions

Instead of revealing a series of analogies to our solar system, the first discoveries exhibited new categories of objects (hot Jupiters and also high eccentricity planets) and new evolutionary mechanisms (migrations, for instance, indicating interaction with the environment). We must now complete the mass and distance distribution, not only for giant planets but also for telluric planets, in order to get sufficient statistics for all planetary systems. At this time, and at this time only, we will be able to identify analogs to our solar system, and understand the importance of the following parameters:

- Number of giant planets
- Position of the giant planets
- Mass and distance distribution of telluric planets
- Potential role of migrations and consequences for the systems
- Origin of migration processes and control parameters
- ...

It also would be interesting to complete this study with objects with masses between those of stars and giant planets, i.e., the brown dwarfs, in order to understand the difference between a giant planet, a brown dwarf or a star during the formation process.

The study of planetary atmospheres is also in its infancy. To date, only a few compounds have been identified in the atmosphere of a single object (HD 209458 b). These few compounds have already allowed the identification of an evaporation phenomenon, which is important for the stability and survival of such an object. The systematic observation of giant planet atmospheres and also those of telluric planets, should allow elucidation of the various environments. This last point will only be attainable if the direct detection and observation of the planets is possible.

From the exobiology point of view the discoveries of exoplanets are very exciting and have led to the proposition of other types of objects, e.g., “ocean planets” as proposed by Léger et al. (2003).

To conclude, it is interesting to see what observational techniques will be possible in the near future, without neglecting the fact that, as recent history has reminded us, we must be prepared for big surprises with these kinds of objects, exactly as we have been with the discovery of hot Jupiters in 1995.

Since the announcement of the first exoplanets discoveries, and particularly 51 Peg b by Mayor and Queloz in 1995, a complete research and study strategy has been developed and progressively implemented. This is the beginning of a new branch of astrophysics and planetology: exoplanetology. The strategy is based on a step-by-step approach to successively search for giant planets by indirect detection methods, search for telluric planets by indirect detection methods, directly characterize giant exoplanets, and directly characterize telluric planets.

6.4.2 Research and Study of Giant Planets

This phase is already under way, as far as radial velocimetry research and planetary transit observations are concerned. It is however interesting to mention what is to be expected from using such methods and to describe future programs.

6.4.2.1 Radial Velocity

The principle of the method is described in Appendix 1.2. As mentioned before, about 4000 mainly F to M spectral type *nearby* stars (within a few tens of parsecs) are being surveyed by several teams around the world. One can thus consider that the search for short-period giant planets (that produce strong radial velocity variations) is rather completed within a radius of 50 pc from the Sun, for solar type or nearby objects. However, new discoveries can be expected at the 50-pc limit. The HARPS instrument program on the 3.6m telescope at ESO reaches an accuracy of 1 m s^{-1} and plans to survey about 1000 objects at distances between 50 and 57 pc. Such measurement accuracy will allow the detection of planets with masses much smaller than those currently observed (for instance a hot Uranus). In addition, a new observation program with the ELODIE instrument (on the 1.93m at OHP) plans to observe 400 new stars at distances of less than 50 parsecs, thus allowing a complete study of the immediate solar neighborhood. The goal of such a study is to get, within a few years, the complete giant planet distribution (mass, distance, eccentricity), to a distance of 5 AU from the parent star for objects in the solar neighborhood, out to a distance of a few tens of parsecs. Even if the method does not allow a determination of the object mass (due to the system inclination), a mathematical deconvolution of the statistic distribution, assuming the equiprobability of inclinations, allows a deduction of the real global mass distribution.

6.4.2.2 Astrometry

The principle of this method is described in Appendix 1.2. Several astrometry programs are planned, from the ground or from space. The astrometric accuracy that will be reachable by ground-based optical and near-infrared interferometers, typically $100\mu\text{as}$ ¹⁵ for the LBT/LINC instrument and from 10 to $50\mu\text{as}$ for the

¹⁵ $1\mu\text{as} = 1\text{ microarcsecond} \sim 5 \times 10^{-12}\text{ rad}$.

VLTI (PRIMA) and Keck-I instruments, should allow the detection of giant planets up to the size of Uranus at the distance of Jupiter.

Up to now, the best astrometric observations are obtained by radio interferometry with continental baselines (VLBI), and with an astrometric accuracy of $100\mu\text{as}$. This method is limited to ten objects that are regularly observed (Lestrade et al. 1999). The future submillimeter interferometer array ALMA, with an astrometric accuracy of about $100\mu\text{as}$, should allow to follow 450 targets in the next decade (Lestrade 2003).

However, it must be mentioned that observing objects at the distance of Jupiter around other stars requires several observation years so that a significant fraction of the orbit is observed. Astrometry certainly will not give complete statistical results for long period objects before several tens of years.

By 2010, future global astrometry missions from space GAIA (ESA) and SIM, should allow by 2010 an astrometric accuracy of several to $1\mu\text{as}$. The clearly announced goal is to try to detect telluric planets around nearby stars. However, a *real* Earth induces a movement that has an amplitude of only $0.6\mu\text{as}$ on a star at 5 parsecs. It therefore appears more than reasonable to assume that this method will mainly find more massive and/or more distant objects. For comparison the astrometric accuracy of the Hipparcos satellite is a few mas.¹⁶

6.4.2.3 Planetary Transits

The principle of this method is described in Appendix 1.2. More than ten ground-based transit follow-up programs are under way. Taking into account the photometric accuracy of ground-based observations this method only concerns giant planets and, practically, such programs have detected very few objects. Up to now 3 OGLE candidates have been confirmed (Udalski et al. 2002), two other are under analysis (Pont et al. 2004), HD 209458 b having been previously detected by radial velocity measurements. The main difficulty with this method is to make the difference between an eclipse due to a small object (planet), a grazing eclipse by a star in a multiple system, or a real eclipse by a faint stellar binary or binary system in the target background. The low luminosity of targets makes the distinction hard.¹⁷

However, we note that transiting planets are in close proximity to their parent star. This allows us to consider the detection and analysis of the variations of the stellar spectrum before and during the transit (this method permitted the detection of several compounds in the atmosphere of HD 209458 b (see Sect. 6.3.2.2), and the light diffused by the planet.

¹⁶ $1\text{ mas} = 1\text{ milliarcsec} \sim 5 \times 10^{-9}\text{ rad}$.

¹⁷ In order to statistically expect several transits, one must observe a crowded region of the sky (several tens of stars per square degrees), i.e., near the galactic plane. One thus typically observes objects several tens of parsecs from us, and with a magnitude of about 15. The spectroscopy of such objects is clearly difficult.

The transit method has greater potential when performed from space.

The methods previously described to detect giant planets and other stochastic methods (gravitational microlensing for instance) should allow a good statistical knowledge of the mass and distance distribution of giant planets in our galaxy, within the next decade. In addition, spectroscopic studies will allow a determination of the atmospheric composition of several planets in the solar neighborhood.

6.4.3 Research and Study of Telluric Planets

Except for the planet candidates detected around pulsars (PSR 1257 +12 for instance), no telluric planet has been detected to date. However, a few methods



Fig. 6.14. Artist's representation of the COROT mission

are potentially adapted to detect these type of objects. Practically, the transit detection method from space is the most promising one in the short-term, and several such programs and missions are currently under way.

The COROT space mission (Fig. 6.14), managed by CNES, is a project that brings together two scientific goals in the same instrument: asteroseismology (observation of stellar vibration modes) and the search for exoplanets by the transit method (Baglin 2003). From 2006, COROT should allow a permanent survey of about 12 000 stars with visible magnitudes between 12 and 16 on a 1.7deg^2 field (Rouan et al. 1999). With a photometric accuracy of a few 10^{-4} , reached by adding over 1 hour photometric measurements collected every 8.5 minutes, COROT will be able to detect Earth size objects. Taking into account the orbit of the satellite (terrestrial polar orbit), each field (and hence each star) will be observed for 5 months at best. The whole program plans to observe 5 fields, i.e., a total of about 60 000 objects.

If we want to observe several transits of the same object, one is limited to the detection of short period objects. On the other hand, and as detailed in Appendix 1.2, the transit probability increases when the distance between the object and the parent star decreases. Figure 6.15 gives the number of transits that COROT should detect for a given field (12 000 stars) and a given type of object, assuming that every star hosts such an object (but that the orbital planes are randomly oriented).

Instrumentally speaking, COROT consists of a 27 cm (equivalent) telescope, a dioptric optical device and a focal plane covered by 4 CCDs, 2 for each program (asteroseismology and exoplanet science). On the exoplanet optical path,

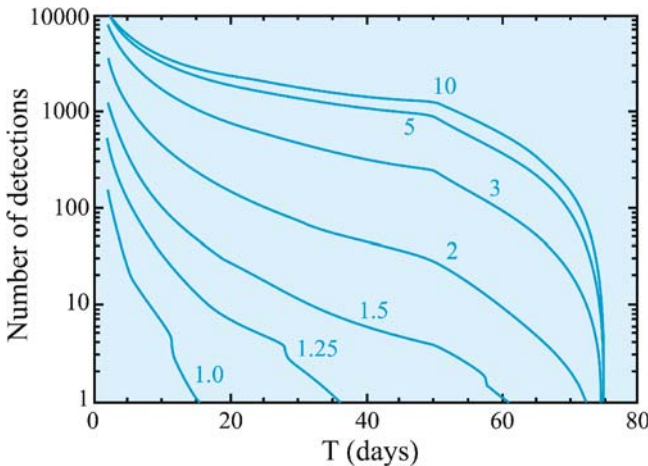


Fig. 6.15. Number of expected detections for one field (12 000 stars) as a function of the object type (period on the abscissa, mass varying from 1 to 10 terrestrial masses on each curve) assuming each star hosts such an object (after Bordé 2003)

a dispersive device (biprism) has been added to get 3-color photometric measurements, over a spectral range from 400 to 900 nm.

The American Kepler Project is a super COROT that aims at detecting real Earth-like planets in the habitable zone of their parent star. In order to reach that goal, Kepler should observe the same field during 3 years and reach a photometric accuracy of a fraction of 10^{-4} . Because the probability of detecting a transiting planet decreases with the distance to the parent star, it is necessary to observe about 100 000 stars simultaneously to expect several Earth-sized planets. Contrary to COROT, Kepler does not include asteroseismology goals. The observatory should be launched in 2008 (Borucki et al. 2003).

The potential of each of these missions, in terms of detectable objects, is shown in Fig. 6.16.

From 2010 to 2012, these various space missions should allow for determination of the mass and distance distribution of telluric exoplanets. In a few cases, it will be possible to observe the transit of several planets in the same planetary system. Unfortunately, in most of the cases, it will not be possible to study the whole planetary system because the previously developed methods to detect giant and telluric planets concern different targets. For instance, for a planetary system seen pole on, in which giant planets will be observable by astrometry, it will not be possible to see transits and thus, to detect potential telluric planets.

To understand the structure of a planetary system, and study the different components, a direct observation of each of the components is needed.

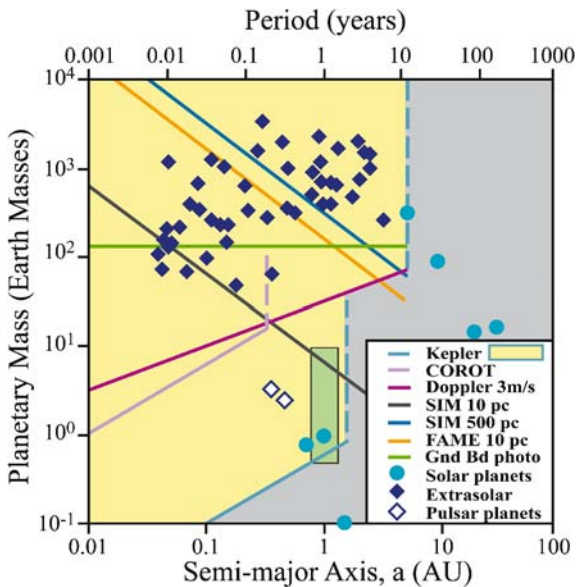


Fig. 6.16. Potential of the future space mission in terms of detectable object types

6.4.4 Characterization of Exoplanets by Direct Detection

In this chapter we will see how we can consider the direct observation of an exoplanet, i.e., physically separating the photons of the object from those of the parent star. The direct observation aims at analyzing each component of the system and particularly its atmosphere to study its composition. We will detail in this section neither the question of planetary atmospheres nor the question of biosignatures. One can find a discussion of these points in the references concerning the Pegasides and in Selsis et al. (2003) for the telluric planets.

6.4.4.1 Posing the Problem

The direct detection of exoplanets is mainly a question of contrast and angular resolution. Figure 6.17 gives the spectral energy distribution of the Sun and the Earth as they would be seen at a distance of 10 pc, as a function of wavelength (Beichman et al. 1999). The Earth's spectrum, as for planets orbiting a star, is made up of the light from the star (the Sun) reflected by the planet (in the case of the Earth around the Sun ($T_{\text{eff}} = 5777\text{K}$), the maximum of $\lambda N(\lambda)$ is at about $1\mu\text{m}$ ¹⁸) and a component linked to the planet's own emission (in the case of the Earth, $T_{\text{eff}} \sim 300\text{K}$ and the maximum of $\lambda N(\lambda)$ is at about $10\mu\text{m}$). The contrast¹⁹ between the Sun and the Earth is thus about 5×10^9 in the visible spectral range (0.5 to $2\mu\text{m}$) and about 7×10^6 in the thermal infrared spectral range (5 to $20\mu\text{m}$). These orders of magnitude are valid for stars with spectral types comparable to that of the Sun. In the visible spectral range the contrast is proportional to the apparent surface of the planet and inversely proportional to the square of the orbit semi-major axis. For instance, the contrast is about 10^9 for a real Jupiter (a planet of the size of Jupiter at the distance of Jupiter, i.e., 5.2 AU). This contrast is reduced to 10^5 for a hot Jupiter situated at 0.05 AU from its star. In the thermal infrared spectral range the contrast depends on the effective temperature of the planet (and hence on its distance from the parent star) and the surface of the planet. Jupiter, thus, even if it has an area 100 times larger than that of the Earth would not be more luminous at $10\mu\text{m}$.

The angular separation ξ (in arcsec) between a planet and its parent star, at a distance L from the observer, is given by the following relation:

$$\xi = \frac{a}{1\text{AU}} \frac{1\text{pc}}{L}. \quad (6.8)$$

An analogy to Earth would thus be observed around a Sun-like star at a distance of 10pc at an angular separation of 0.1 arcsec. The value of ξ should be

¹⁸ The relation between the temperature of a black body T_{eff} (in K) and the wavelength of its maximum photon emission λ_{max} (in meters) is: $\lambda_{\text{max}} = 3.6698 \times 10^{-3}/T_{\text{eff}}$.

¹⁹ Ratio between the stellar and the planetary luminosity at a given wavelength.

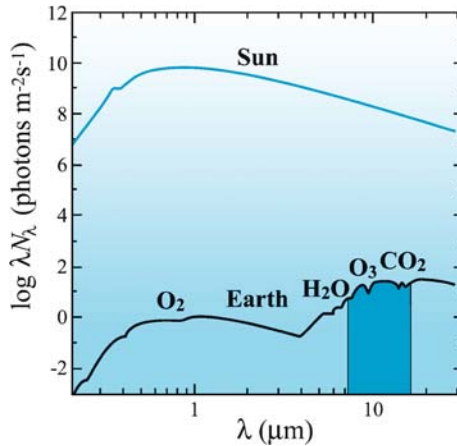


Fig. 6.17. Spectral energy distribution of the Sun and the Earth as seen from a distance of 10 pc as a function of wavelength (after Beichman et al. 1999)

compared to the theoretical angular resolution of a diffraction-limited instrument for which the radius of the diffraction pattern, in the case of a circular pupil, is equal to (in radian, with the relation $1 \text{ arcsec} = 5 \times 10^{-6}$ radian)

$$\theta = 1.22 \frac{\lambda}{D} \quad (6.9)$$

where λ is the observed wavelength and D the telescope diameter. For instance, to observe an Earth-like planet at a distance of 10 pc (at an angular separation of 0.1 arcsec) at $10 \mu\text{m}$ would require a 25 m telescope.

The difficulty in directly observing an exoplanet comes both from the required high angular resolution and high dynamical range. For comparison, observing the Earth around the Sun from a distance of 10 pc is exactly analogous to observing a firefly located in Marseille (France) 30 cm from a lighthouse, with the observer located ... in Paris (France), about 750 km away.

6.4.4.2 The Direct Observation of Exoplanets from the Ground

The direct observation of exoplanets, even giant planets, remains a difficult task from the ground. The sky background intensity (thermal emission of the atmosphere at 300 K) and the atmospheric absorption, mainly due to the presence of water vapor, make observation in the thermal infrared impossible, except in the *N* band ($10.2 \mu\text{m}$, $\Delta\lambda = 5 \mu\text{m}$) and the *Q* band ($21 \mu\text{m}$, $\Delta\lambda = 5 \mu\text{m}$), but where we have to face the problem of variable transmission. In addition, the intensity of the sky background and the thermal background of the instrument require the continuous subtraction of the signal from the background.

In the visible and the near-infrared spectral range the contrast between the star and the planet limits the detection of evolved (cooled) objects located at

more than 2AU from their parent stars. The observation in the visible and in the near-infrared spectral range of hot objects (for instance Pegasides) requires the use of interferometers to reach angular resolutions that allow component separation. The instrumental project GENIE at ESA/ESO (Gondoin et al. 2003) with a main goal of studying stellar zodiacal environments should allow us to observe a few Pegasides (den Hartog et al. 2003) a priori in the L' band ($3.8\mu\text{m}$, $\Delta\lambda = 0.5\mu\text{m}$). Only the observation of massive (big) planets and young objects (several tens of millions of years old) at several astronomical units from their parent stars appears to be reachable using 8m class telescopes. The first image of an exoplanet (2M1207), was obtained using the VLT /NACO instrument (Chauvin et al. 2004). It is a young (eight million years old) massive planet (5 Jupiter masses). The second generation instrument project VLT/Planet finder is a super adaptive optics system coupled to a high dynamics imaging system that should allow for the attainment of this goal (Mouillet et al. 2003).

6.4.4.3 The Direct Observation of Exoplanets from Space

a) Choice of the Spectral Range

The possibility of performing spectroscopy on a detected object is one of the major advantages of space-based observations compared to ground-based. It is then possible to study the composition of the atmospheres (atmospheric gas signatures, CO_2 , NH_3 , CH_4 , etc.) and to potentially look for biosignatures (for instance CO_2 , H_2O , O_2/O_3 under given temperature and pressure conditions).

Once again one can choose to observe in the visible/near-infrared spectral range at the maximum of the diffused solar light or in the thermal infrared spectral range at the emission maximum of the planet. The issue is then to understand what we can effectively observe under given spectral resolution, signal to noise and detection sensitivity conditions. Table 6.2 compares the spectral signatures in both of these spectral ranges.

For the study of terrestrial analogs, Table 6.2 shows that the comparison should be between an observation with a visible observatory ($0.5\text{--}2\mu\text{m}$, spectral resolution of 70 and signal to noise ratio of 25 per spectral bin) and a thermal infrared observatory ($6\text{--}20\mu\text{m}$, spectral resolution of 20 and signal to noise of about 10 per spectral bin). Table 6.3 summarizes the main characteristics of these two spectral ranges.

b) Instrumental Concepts and Associated Projects

As shown in Table 6.3 observations in the visible/near-IR spectral range can be performed from space with a single dish telescope because an object, at the diffraction limit, can be detected at $\lambda = 2\mu\text{m}$ by a 5m class telescope. For comparison, the HST is a 2.4m telescope. Its successor, the JWST (ex-NGST), should have a mirror diameter of about 6m. The main difficulties are the contrast between the star and the planet and the required spectral resolution taking into

Table 6.2. Comparison of atmospheric gaseous signatures in the visible, near-IR and thermal IR spectral ranges. See footnote for the detectability hypotheses (des Marais et al. 2002; Selsis 2002)

Gas	Terrestrial abundance	Spectral signatures (nm)	Visible (0.5–2 μm)		Thermal infrared (6–18 μm)		
			Associated spectral resolution	Detection threshold ²⁰	Spectral signature (μm)	Associated spectral resolution	Detection threshold ²¹
H ₂ O	8000 ppm	1.9, 1.4, 1.1, 0.9, 0.8, 0.7	11, 10, 19, 17, 35, 37	10 ppm	19, 6	3, 3	1 ppm
CO ₂	355 ppm	2, 1.6, 1.2, 1.05	18, 11, 34, 40	100 ppm	15	4	1 ppm
O ₂	~ 21%	1.3, 0.8, 0.7	72, 69, 54	3%	– ²²	–	–
O ₃	6 ppm	0.6	5	1 ppm	9.6	17	1 ppm
CH ₄	2 ppm	1.7, 1.0, 0.9, 0.8, 0.7	10, 20, 32, 29, 57	5 ppm	8, 7.6	6, 13	10 ppm
NO ₂	0.1–1 ppb	–	–	–	6.3	30	100 ppb
NH ₃	0.01 ppb	–	–	–	11, 10, 8.6, 5.8	10, 20, 10, 25	10 ppm

²⁰ Assuming a spectral resolution of about 70 in the 0.5–2 μm spectral range and a signal to noise ratio of about 25 in each spectral bin.

²¹ Assuming a spectral resolution of about 20 in the 6–18 μm spectral range and a signal to noise ratio of about 10 in each spectral bin.

²² O₂ is a homopolar molecule and has no infrared signature.

Table 6.3. Migration time of a planet as a function of its mass and migration mechanism type (Ward 1997)

	Visible	Infrared
Spectral range	0.5–2 μm	6–18 μm
Nb of photons in the spectral range	0.3 $\text{ph m}^{-2} \text{s}^{-1}$	10 $\text{ph m}^{-2} \text{s}^{-1}$
Required spectral resolution	70	20
Required S/N for spectroscopy	25	10
Contrast F_*/F_{pl}	5×10^9	7×10^6
Minimum baseline (diffraction limit)	5m (at 2 μm)	50m (at 20 μm)

account the low photon flux. High contrast observation techniques are under development. These techniques are based on the following:

- Performing phase mask coronagraphs based on the Lyot coronagraph principle (Lyot 1931) but replacing the amplitude mask by a phase mask in the focal plane. The best prototypes of this instrument theoretically allow a reduction in the light from the parent star by a factor of 10^{10} (Rouan et al. 2000). The particularity of the phase mask coronagraph is that it allows for the observation of objects inside the diffraction pattern or in its neighborhood.
- Adapted pupil telescopes in order to limit, or at least to control, the diffraction effects and their orientation. Several pupil shapes have been proposed. For instance, square pupils have their diffraction pattern oriented in two perpendicular directions and allow observation 45° away from these directions, where the image is more or less diffraction-free (Spergel and Kasdin 2001). Gaussian apodised pupils (Nisenson and Papaliolios 2001) have also been proposed.
- A rigorous control of the scattered light. Light scattered by the optical system hampers detection. The reduction of the scattered light to a small fraction of the signal detected requires high-quality optical surfaces (polished typically at 1 Angstroms rms, i.e., about 1 atomic layer).

These instruments, apparently simple on paper, need to be tested and validated in the laboratory. Several prototype coronagraphs are under examination (Riaud et al. 2003). These concepts are the basis of the visible near-IR version of the Terrestrial Planet Finder project (TPF-C), which we will discuss in the following pages.

In the thermal infrared spectral range, and as shown in Table 6.3, the diameter of a single dish telescope that could observe in the 6–18 μm spectral range from space, is too big to be conceivable with present techniques. Interferometry is one solution to this problem. The principle is to combine several small telescopes (each individual telescope does not resolve the planet/star system,

because of diffraction) and to perform aperture synthesis in order to reach the equivalent of a bigger single dish telescope.

One of the most promising concepts that provides both high angular resolution and high dynamic range is known as Bracewell's interferometer (Bracewell 1978). Its principle is described in Appendix 1.2. The DARWIN space mission (Fig. 6.18) and the interferometric version of its American analog TPF-I (Terrestrial Planet Finder) are based on the Bracewell interferometric principle (Fridlund et al. 2000; Beichman et al. 1999).

In one of its design concepts, the DARWIN observatory would consist of $6 \times 1.5\text{m}$ class telescopes located in a regular hexagon. Each telescope, the beam combiner and the communications system will be supported by separate, free-flying, self-controlled satellites equipped with microengines. The satellite positions will be controlled by a metrology system allowing a fixed (but tunable) distance between each component of the array accurate to the millimeter. The



Fig. 6.18. Artist's representation of the DARWIN mission, seen in a six-telescope circular configuration. The *central satellite* is the beam combiner. An eighth satellite, out of the flotilla plane, allows for information transfer to the Earth and the metrology of the array (from an ESA study realized in 1998 by Alcatel Space Industries)

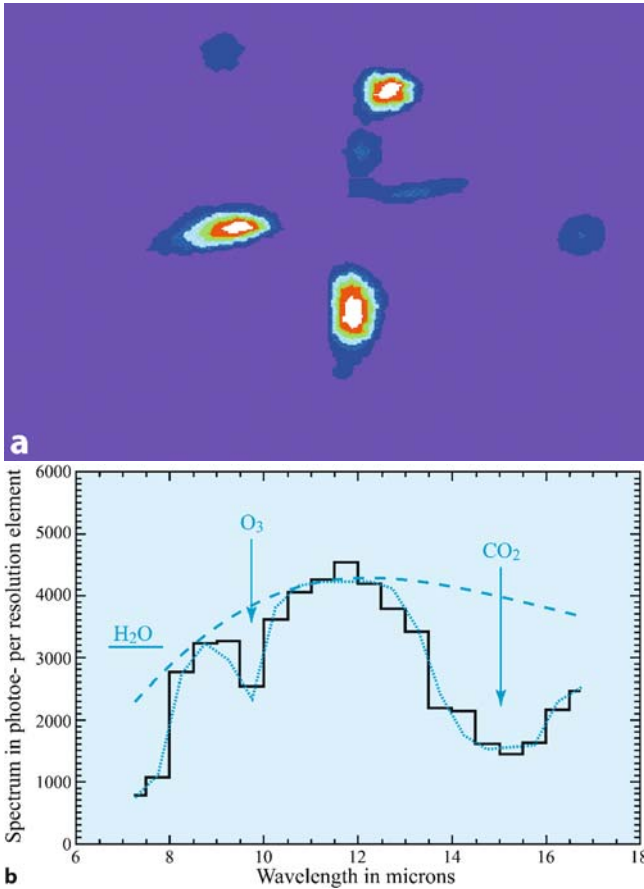


Fig. 6.19. Numerical simulation of the DARWIN nulling interferometric instruments. **a** Solar system map seen from the 01/01/01 pole on from a distance of 10pc from the Sun (*center* of the image), three objects are visible: Venus, the Earth and Mars. **b** Spectrum of the Earth as would be observed by DARWIN; (*histogram*) reconstructed spectrum (observation simulation); (*solid line*) real Earth spectrum at the instrument resolution; (*dashed line*) the 300K black body spectrum. This figure clearly shows that DARWIN/TPF will observe atmospheric gases, even with its limited spectral resolution (20 in this case). In particular, H₂O, CO₂ and O₃ are clearly visible when compared to the 300K black body (after Mennesson and Mariotti 1997)

co-phasing will be performed by delay lines (of a few centimeters). This configuration will allow isolating subarrays inside the global array, which will then be recombined using internal modulation to modulate the signal of the planet with respect to the signal of the star, and to perform a lock-in-like detection, without rotating the whole flotilla as had initially been proposed by Bracewell (Mennesson et al. 2004). Other configurations, with three telescopes for instance, are

also under study. DARWIN/TPF²³ should be launched around 2015 and orbit around the L2 Lagrange point. This point is particularly well-suited to formation flying because of low gravitational gradients. It is also well-suited in terms of the thermal environment. The main DARWIN/TPF instrument will be a nulling interferometer whose aim will be to map planetary systems, built from the modulated chromatic signals from each subarray and the spectra of each planetary system component (Fig. 6.19).

Taking into account the complexity of a nulling interferometric beam combiner, several studies are under way to develop laboratory demonstrators (Ollivier et al. 2001; Brachet et al. 2003; Serabyn 2003a; Flatscher et al. 2003; Barillot et al. 2003) or on the sky demonstrators (Hinz et al. 2003; Serabyn 2003b; Gondoin et al. 2003).

Finally, one cannot conclude this project review without mentioning the potential of hypertelescopes (Labeyrie et al. 2003). In this case an array of several tens of small telescopes is combined thanks to an optical device, called a pupil densifier. This device creates, in a limited field around the pointing axis, the observation conditions of a single dish telescope with the dimensions of the array. The direct observation of planets is then performed thanks to a coronagraph, such as the phase mask coronagraph mentioned before. Assuming a supplementary difficulty level, this concept can probably be considered for the “post DARWIN/TPF” instrument generation.

References

- Allard F., Baraffe I., Chabrier G., Barman T., 2003, Atmospheres of extrasolar planets and brown dwarfs. *ESA-SP*, **539**, 247–252
- Armitage P., Livio M., Lubow S., Pringle J., 2002, Predictions for the frequency and orbital radii of massive extrasolar planets. *MNRAS*, **334**, 248–256
- Baglin A., 2003, COROT: a minisat for pionnier science, asteroseismology and planets finding. *Adv. Space Res.*, **31**, 345–349
- Barillot M., Laramas C., 2003, Alcatel Space involvement in Darwin. *ESA-SP*, **539**, 345–347
- Barman T.S., Hauschildt P.H., Allard F., 2001, Irradiated planets. *Astrophys. J.*, **556**, 885–895
- Beichman C.A., Woolf N.J., Lindensmith C.A., 1999, The Terrestrial Planet Finder: a NASA Origins program to search for habitable planets. *JPL Pub.*, **99-3**
- Black D.C., Matthew M.S. (eds.) 1985, Protostars and Planets 2. University of Arizona Press, Tucson, AZ
- Bordé P., 2003, Détection et caractérisation de planètes extrasolaires par photométrie visible et interférométrie infrarouge à très haute précision. Dissertation, l’Université Pierre et Marie Curie, Paris VI

²³ Taking into account the possible cost of such a mission, ESA and NASA plan a collaboration and maybe a fusion of the DARWIN and TPF-I concepts. This point is under discussion. We will thus mention the project under the common name DARWIN/TPF.

- Borucki W.J., Koch D.G., Lissauer J.J., Basri G.B., Caldwell J.E., Cochran W.D., Dunham E.W., Geary J.C., Latham D.W., Gilliland R.L., Caldwell D.A., Jenkins J.M., Kondo Y., 2003, The Kepler Mission: a wide-field-of-view photometer designed to determine the frequency of Earth-size planets around solar-like stars. *Proc. SPIE*, **4854**, 129–140
- Bouchy F., Pont F., Santos N.C., Melo C., Mayor M., Queloz D., Udry S., 2004, Two new very hot Jupiters among the OGLE transiting candidates. *Astron. Astrophys.*, **421**, L13–L16
- Bracewell R.N., 1978, Detecting non-solar planets by spinning infrared interferometer. *Nature*, **274**, 780–781
- Brachet F., Labèque A., Sekulic P., 2003, Polychromatic laboratory test bench for Darwin/TPF: first results. *ESA-SP*, **539**, 385–387
- Brown T.M., Charbonneau D., Gilliland R.L., Noyes R.W., Burrows A., 2001, Hubble Space Telescope time-series photometry of the transiting planet of HD 209458. *Astrophys. J.*, **552**, 699–709
- Bryden G., Lin D.N.C., 1999, The effects of gap formation and orbital migration on planetary systems. In: Proceedings of 31st AAS/DPS Meeting, Padua, Italy, October 1999
- Campbell B., Walker G.A.H., Yang S., 1988, A search for substellar companions to solar-type stars. *Astrophys. J.*, **331**, 902–921
- Charbonneau D., Brown T.M., Latham D.W., Mayor M., 2000, Detection of planetary transit across a Sun-like star. *Astrophys. J.*, **529**, L45–L48
- Charbonneau D., Brown T.M., Noyes R.W., Gilliland R.L., 2002, Detection of an extrasolar planet atmosphere. *Astrophys. J.*, **568**, 377–384
- Chauvin G., Lagrange A.-M., Dumas C., Zuckerman B., Mouillet D., Song I., Beuzit J.-L., Lowrance P., 2004, A giant candidate near a young brown dwarf. Direct VLT/NACO observations using IR wavefront sensing. *Astron. Astrophys.*, **425**, L29–L32
- Den Hartog R., Absil O., Kaltenecker L., Gondoin P., Wilhelm R., Fridlund M., 2003, Could GENIE detect hot Jupiters? *ESA-SP*, **539**, 399–402
- Des Marais D.J., Harwit M.O., Jucks K.W., Kasting J.F., Lin D.N.C., Lunine J.I., Schneider J., Seager S., Traub W.A., Woolf N.J., 2002, Remote sensing of planetary properties and biosignatures on extrasolar terrestrial planets. *Astrobiology*, **2**, 153–181
- Eggleton P.P., Kiseleva-Eggleton L., 2001, Orbital evolution in binary and triple stars, with an application to SS lacertae. *Astrophys. J.*, **562**, 1012–1030
- Favata F., 2002, The Eddington baseline mission. In: Proceedings of the first Eddington Workshop on stellar structure and habitable planet finding. *ESA-SP*, **485**, 3–10
- Flatscher R., Sodnik Z., Ergenzinger K., Johann U., Vink R., 2003, Darwin nulling interferometer breadboard I: system engineering and measurements. *ESA-SP*, **539**, 283–291
- Fridlund M. et al., 2000, DARWIN, The Infrared Space Interferometer. *ESA-SCI*, **12**
- Goldreich P., Sari R., 2003, Eccentricity evolution for planets in gaseous disks. *Astrophys. J.*, **585**, 1024–1037
- Goldreich P., Tremaine S., 1979, The excitation of density waves at the Lindblad and corotation resonances by an external potential. *Astrophys. J.*, **233**, 857–871
- Goldreich P., Tremaine S., 1980, Disk-satellite interactions. *Astrophys. J.*, **241**, 425–441

- Gondoin P., Absil O., Fridlund C.V.M., Erd C., den Hartog R.H., Rand N., Glinde-
mann A., Koehler B., Wilhelm R., Karlsson A., Labadie L., Mann I., Peacock
A.J., Richichi A., Sodnik Z., Tarengi M., Volonte S., 2003, Darwin Ground-
based European Nulling Interferometry Experiment (GENIE). *Proc. SPIE*, **4838**,
700–711
- Goukenleuque C., Bézard B., Joguet B., Lellouch E., Freedman R., 2000, A radiative
equilibrium model of 51 Peg b. *Icarus*, **143**, 308–323
- Guillot T., Burrows, A., Hubbard W.B., Lunine J.I., Saumon D., 1996, Giant planets
at small orbital distance. *Astrophys. J.*, **459**, L35–L38
- Guillot T., Showman A.P., 2002, Evolution of 51 Pegasus b-like planets. *Astron. As-
trophys.*, **385**, 156–165
- Halbwachs J.-L., Arenou F., Mayor M., Udry S., Queloz D., 2000, Exploring the brown
dwarf desert with Hipparcos. *Astron. Astrophys.*, **355**, 581–594
- Halbwachs J.-L., Mayor M., Udry S., 2005, Statistical properties of exoplanets IV. The
period-eccentricity relations of exoplanets and of binary stars. *Astron. Astrophys.*,
431, 1129–1137
- Henry G.W., Marcy G.W., Butler R.P., Vogt S.S., 2000, A transiting 51 Peg-like planet.
Astrophys. J., **529**, L41–L44
- Hinz P.M., Angel J.R.P., McCarthy D.W., Hoffman W.F., Peng C.Y., 2003, The Large
Binocular Telescope Interferometer. *Proc. SPIE*, **4838**, 108–112
- Horne K., 2003, Status and prospects of planetary transit searches: hot Jupiters ga-
lore. In: Deming D., Seager S. (eds.) *Scientific Frontiers in Research on Extrasolar
Planets. ASP Conf. Ser.*, **294**, 361–370
- Kasai Y., 1962, Secular perturbations of asteroids with high inclination and eccentricity.
Astron. J., **67**, 591–598
- Kuchner M.J., Lecar M., 2002, Halting planet migration in the evacuated centers of
protoplanetary disks. *Astrophys. J.*, **574**, L87–L89
- Labeyrie A., Le Coroller H., Dejonghe J., Martinache F., Borkowski V., Lardière O.,
Koechlin L., 2003, Hypertelescope imaging: from exo-planets to neutron stars. *Proc.
SPIE*, **4852**, 236–247
- Lammer H., Selsis F., Ribas I., Guinan E.F., Bauer S.J., Weiss W.W., 2003, Atmo-
spheric loss of exoplanets resulting from stellar x-ray and extreme ultraviolet heat-
ing. *Astrophys. J.*, **598**, L121–L124
- Lecavelier des Etangs A., Vidal-Madjar A., McConnell J.C., Hébrard G., 2004, Atmo-
spheric escape from hot Jupiters. *Astron. Astrophys.*, **418**, L1–L4
- Léger A., Selsis F., Sotin C., Guillot T., Despois D., Mawet D., Ollivier M., Labèque
A., Valette C., Brachet F., Chazelas B., Lammer H., 2003, A new family of planets?
Ocean-planets. *Icarus*, **169**, 499L–504L
- Léna P. et al., 1996, *Les sciences du ciel*. Ed. Flammarion, Paris
- Lestrade J.-F., Preston R., Jones D., Phillips R., Roger A., Titus M., Rioja M.,
Gabuzda D., 1999, High-precision VLBI astrometry of radio-emitting stars. *As-
tron. Astrophys.*, **344**, 1014–1026
- Lestrade J.-F., 2003, Future astrometry with ALMA to characterize extrasolar planet
orbits. In: Deming D., Seager S. (eds.) *Scientific Frontiers in Research on Extrasolar
Planets. ASP Conf. Ser.*, **294**, 587–590
- Lin D., Bodenheimer P., Richardson D., 1996, Orbital migration of the planetary com-
panion of 51 Pegasi to its present location. *Nature*, **380**, 606–607

- Lyot B., 1931, Photographie de la couronne solaire en dehors des éclipses. *C.R. Acad. Sci. Paris*, **193**, 1169
- Maran S.P., 1992, The Astronomy and Astrophysics Encyclopedia. Van Nostrand Reinhold (New York) and Cambridge University Press, Cambridge
- Marcy G.W., Benitz K.J., 1989, A search for substellar companions to low-mass stars. *Astrophys. J.*, **344**, 441–453
- Marzari F., Weidenschilling S.J., 2002, Eccentric extrasolar planets: the jumping Jupiter model. *Icarus*, **156**, 570–579
- Mayor M., Queloz D., 1995, A Jupiter-mass companion to a solar-type star. *Nature*, **378**, 355–359
- Mayor M., Frei P.-Y., 2001, Les nouveaux mondes du cosmos, A la découverte des exoplanètes. Editions du Seuil, Paris
- Mazeh T., Naef D., Torres G., Latham D.W., Mayor M., Beuzit, J.-L., Brown T.M., Buchhave L., Burnet M., Carney, B.W., Charbonneau D., Drukier G.A., Laird J.B., Pepe F., Perrier C., Queloz D., Santos N.C., Sivan J.-P., Udry S., Zucker S., 2000, The spectroscopic orbit of the planetary companion transiting HD 209458. *Astrophys. J.*, **532**, L55–L58
- Mazeh T., Zucker S., 2003, A possible correlation between mass ratio and period ratio in multiple planetary systems. *Astrophys. J.*, **590**, L115–L117
- Mennesson B., Mariotti J.-M., 1997, Array configurations for a space infrared nulling interferometer dedicated to the search for Earth-like extrasolar planets. *Icarus*, **128**, 202–212
- Mennesson B., Léger A., Ollivier M., 2004, Direct detection and characterization of extrasolar planets: the Mariotti Space Interferometer. *Icarus* in press
- Mouillet D., Fusco T., Lagrange A.-M., Beuzit J.-L., 2003, Planet Finder on the VLT: context, goals and critical specifications for adaptive optics. *EAS Pub. Ser.*, **8**, 193–199
- Murray N., Hansen B., Holman M., Tremaine S., 1998, Migrating planets. *Science*, **278**, 69–72
- Nelson R., Papaloizou J., Masset F., Kley W., 2000, The migration and growth of protoplanets in protostellar discs. *MNRAS*, **318**, 18–36
- Nisenson P., Papaliolios C., 2001, Detection of earth-like planets using apodized telescopes. *Astrophys. J.*, **548**, L201–L205
- Ollivier M., 1999, Contribution à la recherche d'exoplanètes – Coronographie interférentielle pour la mission Darwin. Dissertation, Université de Paris-Sud (Paris-XI)
- Ollivier M., Mariotti J.-M., Léger A., Sékúlic P., Brunaud J., Michel G., 2001, Interferometric coronagraphy for the DARWIN space mission: laboratory demonstration experiment. *Astron. Astrophys.*, **370**, 1128–1136
- Pont F., Bouchy F., Queloz D., Santos N.C., Melo C., Mayor M., Udry S., 2004, The missing link: a 4-day period transiting exoplanet around OGLE-TR-111. *Astron. Astrophys.*, **426**, L15–L18
- Del Popolo A., Gambera M., Ercan N., 2001, Migration of giant planets in planetesimal discs. *MNRAS*, **325**, 1402–1410
- Rasio F.A., Ford E.B., 1996, Dynamical instabilities and the formation of extrasolar planetary systems. *Science*, **274**, 954–956
- Riaud P., Boccaletti A., Baudrand J., Rouan D., 2003, The four-quadrant coronagraph. III. Laboratory performance. *PASP*, **115**, 712–719

- Rosenblatt F., 1971, A two-color photometric method for detection of extra-solar planetary systems. *Icarus*, **14**, 71
- Rouan D., Baglin A., Barge P., Copet E., Deleuil M., Léger A., Schneider J., Toubanc D., Vuillemin A., 1999, Searching for extrasolar planets with the COROT space mission. *Phys. Chem. Earth*, **24**, 567
- Rouan D., Riaud P., Boccaletti A., Clénet Y., Labeyrie A., 2000, The four-quadrant phase-mask coronagraph. I. Principle. *PASP*, **112**, 1479–1486
- Santos N.C., Israelian G., Mayor M., 2001, The metal-rich nature of stars with planets. *Astron. Astrophys.*, **373**, 1019–1034
- Santos N.C., Mayor M., Queloz D., Udry S., 2002, Extrasolar planets. *The Messenger*, **110**, 32–38
- Santos N.C., Israelian G., Mayor M., Rebolo R., Udry S., 2003, Statistical properties of exoplanets: II. Metallicity, orbital parameters, and space velocities. *Astron. Astrophys.*, **398**, 363–376
- Schneider J., 2003, The extrasolar planets encyclopaedia. <http://www.obspm.fr/encycl/encycl.html>. Cited 2003
- Selsis F., 2002, The exo/astrobiology input of a Darwin/TPF mission. *ESA-SP*, **518**
- Selsis F., Léger A., Ollivier M., 2003, Signatures spectroscopiques de vie sur les exoplanètes – les missions Darwin et TPF, dans *Les traces du Vivant*, ouvrage collectif sous la direction de Muriel Gargaux, Presses Universitaires de Bordeaux
- Serabyn E., 2003a, Nulling interferometry progress. *Proc. SPIE*, **4838**, 594–608
- Serabyn E., 2003b, An overview of the Keck Interferometer Nuller. *ESA-SP*, **539**, 91–98
- Showman A.P., Guillot T., 2002, Atmospheric circulation and tides of 51 Pegasus b-like planets. *Astron. Astrophys.*, **385**, 166–180
- Spergel D., Kasdin J., 2001, A shaped pupil coronagraph: a simpler path towards TPF, 199th AAS meeting. *Bull. Am. Astron. Soc.*, **33**, 1431
- Sudarsky D., Burrows A., Hubeny I., 2003, Theoretical spectra and atmospheres of extrasolar giant planets. *Astrophys. J.*, 588, 1121–1148
- Trilling D.E., Benz W., Guillot T., Lunine J.I., Hubbard W.B., Burrows A., 1998, Orbital evolution and migration of giant planets: modeling extrasolar planets. *Astrophys. J.*, **500**, 428–439
- Trilling D.E., Lunine J.I., Benz W., 2002, Orbital migration and the frequency of giant planet formation. *Astron. Astrophys.*, **394**, 241
- Udalski A., OGLE-III Survey Team, 2002, New transits from the OGLE-III survey. *IAU Circ.*, **7925**, 20 June 2002
- Udry S., Mayor M., Santos N.C., 2003, Statistical properties of exoplanets: I. The period distribution: constraints for the migration scenario. *Astron. Astrophys.*, **407**, 369–376
- Vidal-Madjar A., Lecavelier des Etangs A., Désert J.-M., Ballester G.E., Ferlet R., Hébrard G., Mayor M., 2003a, An extended upper atmosphere around the extra-solar planet HD 209458b. *Nature*, **422**, 143–146
- Vidal-Madjar A., Lecavelier des Etangs A., 2003b, Osiris (HD 209458 b), an evaporating planet. *ASP Conf. Ser.* in press
- Vidal-Madjar A., Désert J.-M., Lecavelier des Etangs A., Hébrard G., Ballester G.E., Ehrenreich D., Ferlet R., McConnell J.C., Mayor M., Parkinson M., 2004, Detection of oxygen and carbon in the hydrodynamically escaping atmosphere of the extrasolar planet HD 209458 b. *Astrophys. J.*, **604**, L69–L72
- Ward W., 1997, Protoplanet migration by nebula tides. *Icarus*, **126**, 261

- Wuchterl G., 2000, In: Penny A., Artymowicz P., Lagrange A.-M., Russel S. (eds.) Planetary Systems in the Universe: Observations, Formation and Evolution, IAU Symp. 202, *ASP Conf. Ser.* in press
- Wolszczan A., Frail D., 1992, A planetary system around the millisecond pulsar PSR1257+12. *Nature*, **255**, 145–147
- Zucker S., Mazeh T., 2002, On the mass-period correlation on the extrasolar planets. *Astrophys. J.*, **568**, L113–L116

7 Habitability: the Point of View of an Astronomer

Franck Selsis

7.1 Introduction

Physical and chemical conditions at the surface of the Earth have changed considerably during the last 3.5 Ga.¹ Oxygen for instance became a major component of the atmosphere 2.3 Ga ago. However, during this whole period, and maybe earlier, the Earth was inhabited and thus habitable. This is an observational statement. Theoretically, it is much more difficult to understand what precisely makes the Earth a habitable planet. It is indeed beyond our present knowledge to determine what are the minimum requirements for the emergence of life or for its long-term survival. The approach we will follow in this chapter is to identify some astronomical or geophysical properties that appear to be necessary for the survival of life (as we know it, by default) on a planet or in a planetary system. The reader will keep in mind that these conditions, considered necessary by the astronomer or the geophysicist, are absolutely not sufficient in terms of biology.

To determine a set of necessary conditions for one given type of living species (a human being, for instance) might be possible but the wide diversity of life makes it extremely difficult to infer a more general definition. Thanks to extraordinary adaptation abilities, terrestrial unicellular life forms have colonized some environments we a priori considered extreme or hostile due to their temperature, pressure, radiation level, dryness, or chemical composition. Nonetheless, this adaptation ability seems to face some limits; for instance we have not yet found organisms able to grow at temperatures above 130°C. Considering that these limits are common to any life form in the universe is probably wrong but, in the absence of theoretical limits for life, many consider the limits of terrestrial life as the limits of the habitable universe. In this conservative view, each discovery of a new bacteria thriving beyond these limits makes the habitable universe bigger.

Defining habitability faces another problem: terrestrial life has adapted to dramatic environmental changes. The composition of the atmosphere has changed drastically, from anoxic to oxygen-rich, the pH of the ocean has been modified, the Earth has been submitted to large impacts along its history and even to a heavy bombardment 3.8 to 4.0 Ga ago. Evolution, which was the key

¹ 1 Ga (gigaannum) = 1 Ga (gigayear) = one billion years

to survival to such environmental changes, has resulted in the tremendous diversity and the large tolerance of the present biosphere. However, if we were able to trace the evolution back to its origin we would probably find a much narrower habitable domain. The emergence of life certainly requires more constraining conditions. These conditions are still unknown: we have no precise idea of the conditions on early Earth and no clue about the kind of environment that was associated to the transition from complex organic chemistry to life. Thus, in our search for the mechanisms at the origins of life, we do not know if we should pay more attention to the atmosphere, the oceans, the hydrothermal systems, the surface, the subsurface, the polar ices, the periods following large impacts or volcanic eruptions, lakes, rivers, etc. It becomes even more complicated if we consider the possibility of an extraterrestrial origin of life. This hypothesis, called panspermia, relies on the possible transfer of microbial life from one planet (or other astrophysical object) to another. It has been shown that the transfer from Mars to the Earth was not only possible but also difficult to avoid if life appeared early in Mars' history. Five percent of the rocks ejected from Mars end on Earth, which explains the presence on Earth of numerous Martian meteorites. Bacteria can survive the impact necessary to eject some Martian rocks to space as well as the heating during the entry into Earth's atmosphere and the impact on Earth's surface (Mastrapa et al. 2001). Inside the ejected rock, they can survive the flux of cosmic rays (Horneck et al. 1994). Interstellar panspermia, from one stellar system to another, has also been proposed as a possible mechanism for the origin of life on Earth (Napier 2004). At the moment, it is difficult to address scientifically the role of panspermia in the origin of terrestrial life because of the lack of means to test the hypothesis. Most of the scientists consider that suggesting another cradle of life than Earth is unfruitful, as it does not help us to narrow our search for the prebiotic environment associated with the appearance of life. However, it is important to keep in mind that science in its present state cannot rule out nor confirm the panspermia hypothesis. In our discussion we will consider that habitability is the property of some astrophysical object to host life, whatever its origin and wherever it comes from.

A practical approach, followed by most of the studies addressing habitability in an astrophysical context, consists in defining *habitable* no longer as *where life can exist* but as *where liquid water can exist*. Of course, if biologists agree on the fact that life requires liquid water, life obviously requires much more. We could add more criteria, like the presence of the main chemical elements used by life, such as carbon and nitrogen, but we are anyway far from being able to define a minimum set of requirements for the simplest life form. The presence of liquid water is usually considered as requisite for life and the physical and chemical conditions at which it is stable can be traced by astronomical means. In the following, we will thus use this operational definition and identify the regions where liquid water, and thus life as we know it, cannot exist; the remaining (small) fraction of the universe where liquid water can reasonably be searched for being the *habitable* universe.

One of the goals of astrobiology is to search for signatures of life on other planets (or satellites) in the solar system but also beyond, around other stars. In this context, the question the astronomer has to address, as an observer, is where to search? When we look beyond our own system, we discover an incredible diversity of systems, depending of course on how far we can explore our nearby universe. With the first space telescopes dedicated to the search for habitable and inhabited worlds around other stars (Darwin: Léger et al. 1996; Selsis et al. 2003; TPF-C: Beichman et al. 1999) we will not be able to observe all the accessible stars. This is due to the long integration times required for the planet detection (hours to days) and for the spectral characterization of the planets found (weeks). To select its future targets, the astronomer will thus use the liquid water criterion, adapted to its observing capabilities. What we look for is a planet where the biosphere is active and widespread enough to modify the observable planetary properties (which can be the atmospheric or surface composition). For instance, a planet that hosts reduced ecosystems confined in the subsurface, with a metabolic activity too low to affect the superficial environment would not exhibit signs of life to the remote observer. This is the problem we face with Mars or Europa: life might be there, thriving in the subsurface, but we cannot conclude remotely on the presence or absence of biological activity. For this reason, we will consider an even more restrictive definition: a planet is considered habitable when liquid water can exist at its surface, which implies, among other things, an atmosphere. This definition can be seen as a compromise between biological and observational approaches.

We should also note that the presence of liquid water at the surface is necessary to allow photosynthesis. Indeed, photosynthetic organisms need both light and liquid water and can thus only grow at the surface or within the upper layers of water. Some scientists have suggested that complex multicellular life (animals and plants) require an oxygen-rich atmosphere (Catling et al. 2005). As oxygenic photosynthesis is responsible for the build-up of atmospheric oxygen, one could argue that a planet able to sustain liquid water at its surface is also an habitable planet for complex life requiring atmospheric oxygen (see Sleep 2005 for a comment on that point).

7.2 The Circumstellar Habitable Zone

The habitable zone (HZ) around a star was defined by Hart (1979) as the region where liquid water can exist at the surface of a planet. In other words, it is the range of orbital distance where a planetary surface can be stabilized by climatic and geophysical processes within a temperature and pressure domain compatible with the existence of liquid water (273–647K; 0.006–220bar). The surface temperature of a planet depends on the incident stellar energy and its spectral distribution, the radiative properties of the atmosphere (absorption, scattering, emission) and the reflection properties (albedo) of the surface and the clouds. Kasting (1988) and Kasting et al. (1993) revisited Hart's concept of the HZ by

applying realistic atmospheric models and considering the geochemical cycles of atmospheric constituents (mainly CO_2). They considered the Sun-Earth system and artificially changed the orbital distance of the Earth to study the impact on the climate. By doing this, they determined the minimum and maximum distances at which Earth's ocean is, respectively, totally vaporized and totally frozen. They also estimated the extension of the HZ around other types of stars.

7.2.1 The Inner Limit of the Habitable Zone

Earth's atmosphere contains about 10^6 kg of water vapor, which represents 0.001% of the ocean mass. This ratio between vapor and liquid water is extremely sensitive to the surface temperature. Decreasing the orbital distance of the Earth would increase the abundance of atmospheric water vapor. As a consequence, the atmosphere would become more opaque at infrared wavelengths, making the greenhouse effect more efficient. However, water vapor also makes more efficient the transport of heat by convection from the surface to the lower atmosphere. The warm water vapor vaporized at the surface goes upward and carries latent heat that is released higher in the atmosphere when vapor reaches cooler atmospheric layers and condenses again. This feedback is extremely important as it allows the climate system to find an equilibrium by limiting the increase of the surface temperature in response to an increase of the incoming solar energy. Without this feedback, liquid water would not be stable at the surface of the Earth. Radiative-convective models (including radiative transfer, thermodynamics and convection) can compute the vertical temperature and water vapor profiles of the atmosphere (Kasting 1988, Fig. 7.1). For surface temperatures above 360 K (corresponding to a distance of 0.95 AU from the present Sun and about 1 bar of H_2O at the surface), the outgoing infrared radiation reaches a maximum value. This is due to the fact that, first, the optical depth of the atmosphere becomes much larger than unity within the entire thermal infrared domain. Second, the part of the atmosphere that is radiating into space (where the optical depth is unity) becomes isothermal with altitude and is no longer affected by changes of the surface temperature. Once this threshold is reached, what is the consequence of an increase of the incoming radiation? The excess of radiation that cannot be re-emitted by the planet is then converted into latent heat of water vapor. At 0.95 AU, if this excess represents 1% of the incoming solar energy, the terrestrial oceans could be vaporized in 50 000 years. At that stage, the increase of the albedo associated with the build-up of a dense H_2O atmosphere (10 meters of water = 1 bar) and its Rayleigh back-scattering to space, is the only process protecting the ocean against a complete vaporization. However, this protection does not prevent the surface from reaching the critical temperature of water (647 K) and thus the complete vaporization of the water reservoir, at a distance of about 0.84 AU of the present Sun (equivalent to 1.4 S, S being the present solar energy flux at 1 AU). Once the whole ocean is vaporized (which would produce a 270 bar H_2O atmosphere in the case of the

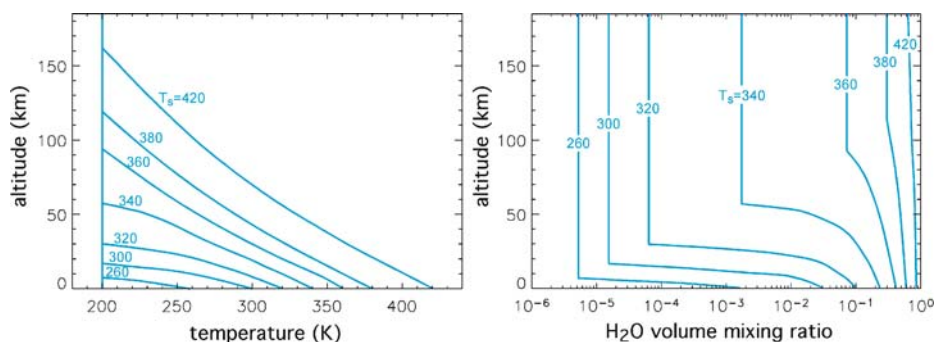


Fig. 7.1. Temperature and water vapor atmospheric profiles as a function of the mean surface temperature (T_S). For high values of T_S , but still compatible with a surface liquid ocean ($T_S < 647\text{K}$), H₂O becomes the main atmospheric species. The partial pressure of H₂O reaches 1 bar at $T_S \approx 400\text{K}$. When T_S rises above 350 K, the atmospheric escape of hydrogen is no longer diffusion-limited, and an amount of hydrogen equivalent to the content of one terrestrial ocean can be lost in less than 1 Ga. See Fig. 7.2 for the relation between T_S and the orbital distance. Modeling and figures by Kasting (1988)

Earth), the surface temperature increases dramatically to temperatures above 1500 K in a phase called runaway greenhouse. At such high temperatures, a new radiative equilibrium is reached, as the absorbed stellar radiation can then be balanced by the fraction of the surface thermal radiation that is emitted in the near infrared and the visible through some spectral atmospheric windows.

As seen above, the orbital distance at which runaway greenhouse was estimated at 0.84 AU from the present Sun. This value should however be considered as an indicative estimate more than as a robust and precise value. Indeed, the models used to determine this limit do not include clouds. High altitude water clouds can be expected in a dense H₂O atmosphere, resulting in a significantly higher albedo. For instance, clouds on Venus are the main contributors to the 70% albedo of the planet, which would not exceed 30% in the absence of clouds. An albedo of 70% could sustain a liquid ocean at orbital distances as close as 0.5 AU from the present Sun.

It should be noted also that within the HZ, defined as above, ocean temperatures can exceed by far the highest temperatures tolerated by known terrestrial hyperthermophiles ($\sim 113^\circ\text{C}$, which is the mean surface temperature that would have a cloud-free Earth at 0.93 AU).

Within the HZ but close to its inner limit, habitability is also threatened by the loss of hydrogen to space. At orbital distances closer than 0.95 AU (FPS²), H₂O becomes the main atmospheric compound, even at high altitudes where

² FPS: from the present Sun. Because the solar luminosity increases with time, we need to use the present solar luminosity as a reference value in order to express the limits of the HZ as orbital distances.

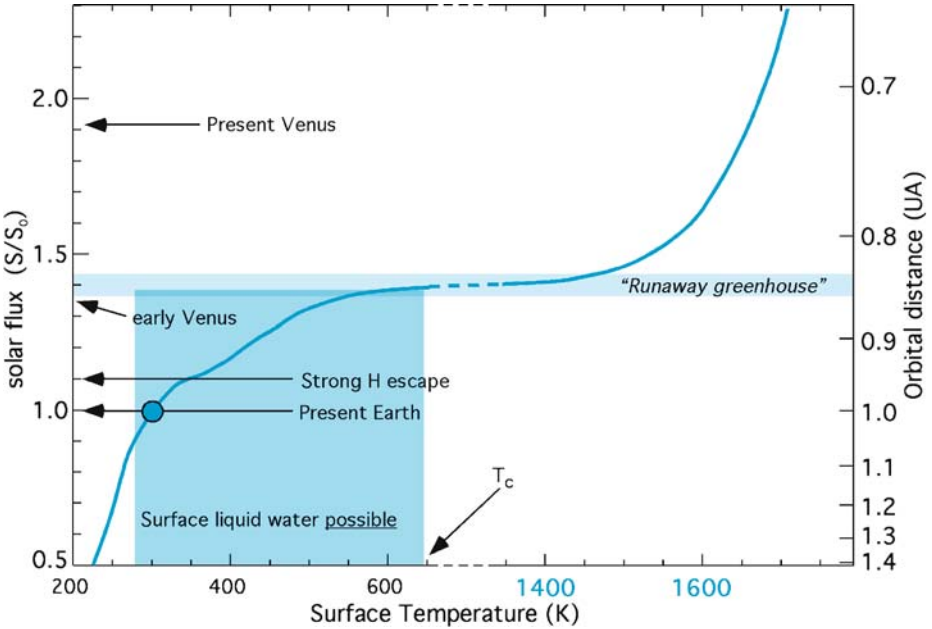


Fig. 7.2. Runaway greenhouse. This graph shows the relation between the solar flux at the top of the atmosphere and the mean surface temperature T_S . At $T_S \approx 550\text{K}$ (equivalent to 1.4AU from the present Sun), the outgoing infrared radiation at the top of the atmosphere reaches a threshold: any small increase of the incoming radiation (or a decrease of the orbital distance) triggers a runaway of the greenhouse effect vaporizing the whole water reservoir and heating the surface up to more than 1500K. This high surface temperature allows the planet to find a new radiative equilibrium by cooling its surface through the atmospheric spectral windows in the near-infrared and the visible. The solar fluxes received by Venus now and 4Ga ago are indicated, as well as the flux resulting in a rapid H escape and temperature/flux region where surface liquid water is stable. This plot assumes that the water reservoir of the planet is larger than any amount of atmospheric water vapor required at equilibrium. This is the case for the Earth but might not be for planets with a lower water content for which the whole reservoir might be vaporized before reaching the critical temperature of 647K. The correspondence between the flux and the distance to the *present* Sun is given on the *right Y-axis*. Note the cut in the *X-axis* between 600 and 1400K. Figure adapted from Kasting (1988)

H_2O is photolysed into O and H by the stellar UV radiation. On Earth, the loss of hydrogen to space is limited by the flux of H-bearing species from the low to the high atmosphere. The low temperature at the tropopause (at an altitude of about 10km) produces the condensation of the ascending water, keeping the abundance of hydrogen at a low level in the stratosphere and above. As shown on Fig. 7.1, the efficiency of this “cold trap” decreases with the orbital distance and when H_2O becomes a major constituent of the high atmosphere, the escape

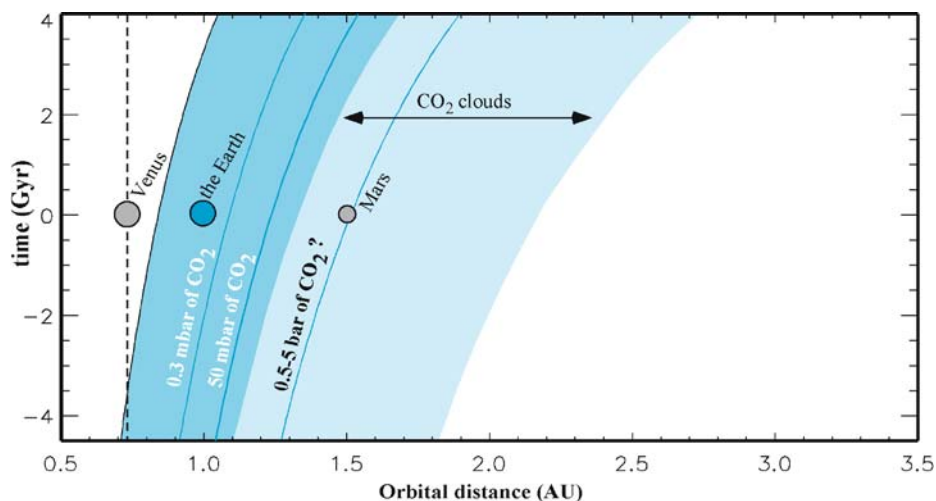


Fig. 7.3. Evolution of the circumstellar habitable zone (HZ). Planetary habitability requires a greenhouse warming increasing with the orbital distance. This figure gives the CO_2 level that insures a mean surface temperature of 0°C at various distances from the Sun and as a function of time. On an Earth-like planet inside the HZ (*whole blue area*), with tectonics and a carbonate-silicate cycle, this minimum CO_2 level should be sustained by geochemical and climatic feedbacks (Walker et al. 1977; Kasting et al. 1993). Closer than the inner edge of the HZ, water is vaporized as seen on Figs. 7.1 and 7.2. Farther than the outer edge of the HZ, the CO_2 greenhouse warming is insufficient to maintain T_S above 273K. The position of this outer edge is not precisely determined because of the complex role of CO_2 clouds (Forget and Pierrehumbert 1997; Mischna et al. 2000) that form in the region represented in *light blue*. For the same reason, the level of CO_2 that would be necessary to make present Mars habitable is uncertain: lowest estimates taking into account CO_2 clouds are about 500 mbar (personal communication with Forget, 2005) while a cloud-free atmosphere requires 5 bars. Note that due to the evolution of Sun's luminosity, Venus was probably habitable during the first hundred million years of its history

rate of H is no longer diffusion-limited and depends only on the temperature of the exosphere (the altitude at which collisions are negligible). The exospheric temperature of the Earth is about 1200K. At such temperatures, an Earth at 0.95AU (FPS) would lose its water content in about 1Ga. This might have been the fate of Venus in its early history. During the first hundred million years after its formation, Venus could have sustained a hot ocean, as shown in Figs. 7.2 and 7.3. With the increase of the solar luminosity, the hydrogen content of the water would have been lost, maybe even before experiencing the runaway greenhouse, depending on the initial amount of water on Venus. This illustrates another constraint on planetary habitability: the water content. First of all, as the critical point of water corresponds to a pressure of 220bar, a water reservoir equivalent to a vapor pressure lower than this value would be fully vaporized

before reaching $T_c = 647\text{K}$. Vaporizing the terrestrial ocean would result in a 270bar atmosphere and would thus require heating the surface above T_c . A less massive reservoir could be fully vaporized at lower temperature. For instance, a 100m deep ocean at the surface of an Earth-like planet (without clouds) would be vaporized at 450K, at an orbital distance of about 0.9AU (FPS). The second reason why the initial amount of water is important is its erosion through H escape. However, the estimates that have been done to quantify the loss rate do not include the fact that an upper atmosphere dominated by H_2O would be colder than it is on present Earth. The reason is that H_2O is a very efficient radiator in infrared than O_2 . The resulting cooling could yield an exospheric temperature closer to 300K than to 1000K and drastically limit the loss of H to space.

7.2.2 The Outer Limit of the HZ (or How to Warm Early Mars?)

On Earth, the main greenhouse gas is water vapor. However, as the partial pressure of H_2O is thermodynamically buffered by the equilibrium with the liquid water reservoir, the water vapor abundance as well as the surface temperature T_S , are determined by the distance to the Sun and by the abundance of the other main greenhouse gas: CO_2 . On an Earth-like planet, with emerged continents and active volcanism, the CO_2 level should also depend on the orbital distance. This arises from the carbonate-silicate cycle and its temperature dependency. Atmospheric CO_2 is released by volcanoes and precipitates into carbonates in the presence of liquid water and land weathering. If the CO_2 level is too low to maintain the surface temperature above 0°C , carbonates are no longer formed and CO_2 builds-up in the atmosphere leading to an increase of the greenhouse warming. Provided that the reservoir and production of CO_2 are large enough, this mechanism inevitably leads to the melting of the surface water and thus to habitable conditions. On Earth, volcanoes release the equivalent of the present CO_2 level (0.3mbar) in about 10 000 years. Inversely, if the volcanic release of CO_2 exceeds its loss into carbonates, the level of CO_2 and the greenhouse warming increase, enhancing the precipitation, the weathering and thus the formation of carbonates. This negative feedback limits the build-up of atmospheric CO_2 . Walker et al. (1981) described in detail this climate/ CO_2 stabilization on which relies the definition of the HZ.

As a consequence of this regulation process, if the Earth had a larger orbital distance, the CO_2 level would stabilize at a higher value. As shown on Fig. 7.3, a temperature of 0°C at 1.2AU (FPS) requires about 50mbars of CO_2 (170 times the present atmospheric level). Estimating the distance of the outer edge of the HZ thus consists of determining at which distance T_S is below 0°C whatever the CO_2 level. Above a CO_2 partial pressure of about 8 bars, increasing the CO_2 level lowers the surface warming. This is due to the Rayleigh scattering by CO_2 molecules: if the atmosphere is too dense, it scatters more stellar energy back to space than thermal energy back to the surface. This threshold of greenhouse inefficiency corresponds to an orbital distance of 1.65AU (FPS) or 0.36S. One can

note that present-day Mars is inside this limit and some authors have suggested the possible terraforming of the planet (McKay et al. 1991). If the Martian surface is not habitable now, this is probably the consequence of the small size of the planet, which has two negative effects on habitability. First, the internal heat flux decreased fast and is now too low to maintain a globally active volcanism that could build-up a denser CO₂ atmosphere. Second, the low gravity (and the absence of magnetic field) enhances the efficiency of atmospheric loss to space. A more massive planet could have been habitable at the orbital distance of Mars (the small size of Mars may be a consequence of the proximity of Jupiter: we will discuss further the gravitational influence of the giant planet on the formation and the composition of habitable planets). Some geological evidence shows that large amounts of stagnant and flowing liquid water were present at the surface of Mars 3.8Ga ago, during periods that are long enough to imply a warmer climate. According to our definition, early Mars was thus habitable. However, 3.8Ga ago, Mars received only 0.32S, which is below the lower limit of 0.36S. It is thus extremely difficult to explain how the average surface temperature was maintained above 0°C. Once again it may be the microphysics of clouds that solve this apparent paradox.

Beyond an orbital distance of about 1.35AU (FPS), atmospheric CO₂ condenses in the middle atmosphere. The CO₂ ice clouds increase the albedo and thus cool the surface. If cooling was the only effect of CO₂ clouds, the condensation of CO₂ would mark the edge of the outer habitable zone. However, the CO₂ ice grains that form in the atmosphere have a typical size of about 10 microns and thus scatter the infrared radiation back to the surface, producing an additional greenhouse warming. Forget and Pierrehumbert (1997) showed that this surface warming exceeds the cooling due to the increase of the albedo and that CO₂ clouds can sustain a surface temperature above 0°C as far as 2.2AU (FPS), equivalent to 0.21S. Irradiated by 0.32S, early Mars could thus have been habitable thanks to the radiative effects of CO₂ clouds. However, the minimum level of CO₂ required to warm early Mars with CO₂ clouds still needs to be determined. If this level is high (\gg 500mbar), it might be difficult to explain why no carbonate has been detected on Mars yet. The high obliquities allowed by Mars' chaotic behaviour could have played a role in the existence of warm and wet regions on early Mars. Considering our present understanding of climate, it seems reasonable to consider that a conservative distance for the outer edge of the HZ corresponds to the case of early Mars (1.5AU (FPS), 0.32S). The maximum extension of the HZ, relevant for larger planets able to sustain a denser CO₂ atmosphere, could not exceed 2.2AU (FPS) (0.21S).

The presence of greenhouse gases other than H₂O and CO₂ could expand the HZ or allow the habitability at lower CO₂ levels. Studies of ancient paleosols (Rye et al. 1995; Hessler et al. 2004) suggest that the level of CO₂ between about 3 and 2.3Ga was lower than the minimum level required to warm the surface of the Earth above 0°C. At these epochs, it is likely that the methane produced by primitive methanogenic bacteria was responsible for part of the

greenhouse warming (Pavlov et al. 2000; Selsis 2000). Back then, the atmosphere contained negligible amounts of O_2 , which allowed CH_4 to build up. The present biogenic emission of methane could have resulted in more than 100 times its present atmospheric level. Moreover, in the absence of O_2 and the presence of widespread sources of H_2 , anaerobic methanogens could grow nearly everywhere on Earth, resulting in a higher CH_4 production. Thus it appears that methane was an important atmospheric compound, which could explain why the CO_2 level was lower than expected from a carbonate-silicate regulation. The Earth could have been quite warm during this golden age of the methanogens but two mechanisms should have limited significantly the rise of the mean surface temperature. First, the photolysis of CH_4 in an anoxic atmosphere initiates the formation of organic haze, like in the Titan atmosphere, increasing the albedo. Second, the absorption of the incoming solar radiation, in the visible and near-infrared ranges, becomes important at high levels of CH_4 . The energy deposited this way in the middle atmosphere produces a stratospheric warming (like in the Neptune atmosphere) but does not reach the surface. As a consequence, the greenhouse warming by CH_4 reaches an optimum for levels of methane on the order of 10 mbars and decreases if the CH_4 level exceeds this value.

Can methane make a planet habitable in the absence of life? CH_4 is a well-known compound of the atmospheres of the outer solar system, but there is no obvious pathway for an efficient and continuous production of CH_4 on terrestrial planets closer to the Sun. On present-day Earth, the main source of methane is life, but a very small fraction of it is released abiotically in hydrothermal systems. The mechanism for the production of CH_4 (and organics in general) through hydrothermalism is complex and catalyzed by various minerals. It can be roughly summarized as follows: at the pressure-temperature ranges found in some hydrothermal systems, water oxidizes iron and releases H_2 , which reacts with CO_2 to form methane. In theory, such process could release more CH_4 than on present-day Earth, provided that physical and chemical conditions are gathered. There is however no known geochemical regulation process that could maintain the CH_4 level at a value corresponding to a mean surface temperature slightly above $0^\circ C$. Thus, transient but not long-term habitability can be provided by CH_4 when its geothermal production fits with the atmospheric level required for habitability. Could methane warm early Mars as it probably did, later, on early Earth? A cloud-free atmosphere, made of CO_2 , CH_4 and H_2O , seems to be unable to warm early Mars because of the stratospheric absorption by methane described earlier (personal communication from Kasting, 2005). On an early Mars warmed by CO_2 clouds, CH_4 could have contributed to the surface warming. However, if the CH_4 level was too high, it would have not only decreased the greenhouse effect by absorbing the solar energy higher in the atmosphere, but the resulting stratospheric warming could have prevented the formation of CO_2 clouds. Methane thus seems to have played a minor role in the climate of early Mars.

As a conclusion, the absence of a simple model to explain the early warm climate of Mars questions the presently published distances of the outer limit of the HZ. We can however consider the 0.36S (1.67AU – FPS) limit as a pessimistic one. Indeed, early Mars seems to have been habitable while receiving less solar energy and additional greenhouse gases could extend this external limit. The concept of HZ should not rely only on the radiative properties of CO₂ but should include the other atmospheric compounds able to contribute to the surface warming. Nevertheless, CO₂ is the only greenhouse gas for which we know a geochemical mechanism that stabilizes its level and that is compatible with an O₂-rich atmosphere. Most of the other greenhouse gases are reduced compounds (CH₄, NH₃, N₂O), whose abundances are very limited when the level of O₂ is high. Therefore, the present definition of the HZ is a relevant definition of the region where O₂ build-up can occur (Selsis 2002).

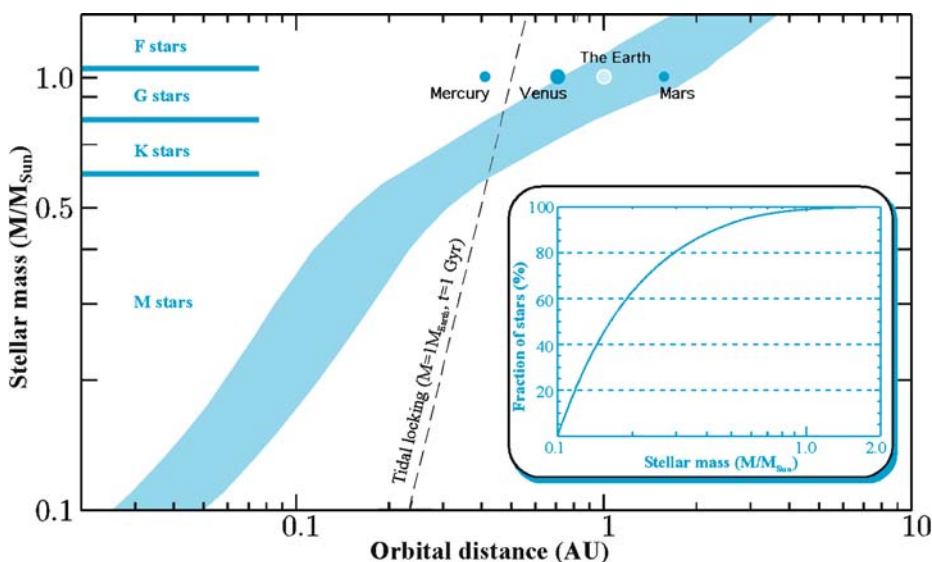


Fig. 7.4. Limits of the continuously habitable zone (CHZ) as a function of the stellar mass. This graph shows the limits of the circumstellar region that remains habitable during more than 1Ga, for stars with a mass between 0.1 and $2M_{\text{Sun}}$. Stars more massive than $2M_{\text{Sun}}$ do not live more than 1Ga and have thus a lower astrobiological interest. The *inserted plot* gives the mass distribution of stars in our galaxy, where we can note that more than 95% of the stars in this mass interval have a mass below $0.6M_{\text{Sun}}$ (M stars). The *dashed line* indicates the orbital distance at which the rotation of a $1M_{\text{Earth}}$ planet (on a circular orbit) would be tidally synchronized in less than 1Ga. The habitability of synchronously rotating planets is an important issue (Joshi 2003), especially considering that most habitable planets may evolve into such a state. Figure from Selsis and Grießmeier (stellar luminosities from Baraffe et al. 1998)

7.2.3 Continuously Habitable Zone

The solar luminosity has been increasing since the formation of the solar system and will continue to rise until the phase of red giant (Gough 1981; Baraffe et al. 1998).

As luminosity grows, the limits of the HZ are pushed away from the Sun, as seen in Fig. 7.3. For this reason, Venus was probably inside the HZ during its first hundred million years but is no longer inside it. It is then logical to define a “continuously habitable zone” (CHZ) that remains habitable during a time t . The choice of t is not obvious, especially if we want to generalize the definition to any kind of star that lives more or less than the Sun. Stars twice as massive as the Sun live only 1 Ga while the lifetime of stars half as massive as the Sun live (in theory) more than 80 Ga. The limits of the CHZ are usually estimated for $t = 1$ or 5 Ga. Figure 7.4 shows the limits for $t = 1$ Ga obtained for different types of stars, and determined using the model for stellar evolution by Baraffe et al. (1998).

7.3 Habitability Around Other Stars

Stars can be characterized by their mass, age, metallicity (i.e., the abundance of elements heavier than He) and by their membership to a multiple stellar system. They can also be characterized by their spectral type, which is related to their temperature and determined by their mass and age. For the nearby stars, we can now also distinguish the stars with or without known planetary systems.

Knowing the limits of the HZ estimated for the present Sun, can we infer the limits for any other star knowing its luminosity L_* ? This can be done for stars having a spectral type (and thus an effective temperature) close to Sun’s, by using the following formula:

$$D_* = D_{\text{sol}}(L_*/L_{\text{sol}})^{1/2} \quad (7.1)$$

where D_* is the distance at which the energy flux from star with a luminosity L_* is equal to the energy flux received at a distance D_{sol} from the present Sun (which is L_{sol}). By replacing D_{sol} by the limits of the HZ for the present Sun, one can infer a very good approximation of the limits of the HZ for another star or for the Sun at another age. In the case of stars whose spectral type (or effective temperature) is too different, this approximation is no longer reliable. This is due to the fact that the albedo of a planet depends strongly on the spectral distribution of the stellar energy received by the planet. Earth’s present albedo is about 0.3 but this is not an intrinsic value: the albedo would be different if, for the same atmospheric composition, an F or M star instead of a G star, would irradiate the Earth. Hot (F) stars emit a larger fraction of their energy in the UV range while cool (M) star luminosity is shifted towards the infrared. As Rayleigh scattering varies as λ^{-4} , the fraction of back-scattered energy, and thus

the albedo, increases with the stellar temperature. Moreover, the absorption of the incoming light by the atmospheric compounds implied by habitability (H_2O , CO_2) is low in the visible but high in the near-infrared. As a consequence, the albedo decreases with the effective temperature of the star. This effect was studied by Kasting et al. (1993) and is included in the calculations made to produce Fig. 7.4.

The definition of the HZ, we gave earlier only relies on the stellar luminosity and its spectral distribution. However, nothing implies that planets exist inside the HZ of a given star, or that planets inside the HZ are necessarily habitable. We already mentioned the importance of the planetary mass for habitability but, as we will see, many more factors have to be taken into account.

The Sun is a G-type star, with a lifetime of about 10 Ga. More massive stars live less: stars above $2M_{\text{Sol}}$ die before reaching 1 Ga. As a consequence, massive stars could only host ephemeral biospheres that have probably not enough time to develop and modify their planetary environment in a way that could be remotely detected. These massive stars represent however a very small minority in the stellar population of the galaxy. In the $0.1\text{--}2M_{\text{Sol}}$ range, 95% of the stars have a mass below $0.6M_{\text{Sol}}$. The habitability of planets orbiting these stars (M-type stars) is thus a very important issue. Because of their low luminosity, the HZ is very close to the central star and potentially habitable planets are submitted to strong tidal effects. In the HZ of M stars, the rotation of terrestrial planets on circular orbits is rapidly synchronized with the revolution period and the planet eventually always shows the same face to its star. Is such a planet habitable? The question is still opened. Joshi et al. (1997) showed that a thick CO_2 atmosphere (on the order of 1 bar) is required to prevent the atmosphere from condensing on the cold dark side. It is however not known if a carbonate-silicate cycle and climatic feedbacks could stabilize the level of CO_2 at such high values. Taking into account the heat transported by oceanic streams and water vapor could result in a lower minimum level of CO_2 (Heath et al. 1999; Joshi 2003). Planets in the HZ of M stars have to face another problem: the activity of their central star and the consequent atmospheric losses. While the present Sun emits only 10^{-6} of its total luminosity in this wavelength range, this ratio is 1000 times higher in the case of M stars. This strong XUV luminosity is associated with powerful particle winds and coronal ejections that erode the planetary atmospheres in the HZ. The protection against this particle flow by the magnetic field is probably less efficient inside the HZ of M stars where the planet rotation is slowed down by tidal interactions, weakening the internal dynamo at the origin of the magnetic field. To quantify the continuous mass loss experienced by planets around M stars is a complicated task and it is not clear yet how affected the habitability is. It seems however clear that if inhabited planets do exist around M stars, their atmosphere and climate as well as the ecosystems they host, must be radically different from what is found on Earth.

Most of the stars in our Galaxy belong to binary or multiple systems. The presence of one or several stellar companions can result in gravitational instabilities in the HZ of a given star of the group. It may thus be possible that a large fraction of the stars cannot form habitable planets or host them during their whole life. Another factor that should influence the habitability of a stellar system is its metallicity (i.e., the abundance of elements heavier than He), which is inherited from the interstellar medium where the star formed. As terrestrial planets consist in these heavier elements, they cannot form below a minimum value of the metallicity. On the other hand, observation seems to show that stars with high metallicity tend to host close-in giant planets, which are, in most cases, incompatible with the presence of terrestrial planets in the HZ. Using these constraints, Lineweaver et al. (2004) defined a galactic habitable zone, which is a ring between 7 and 9kpc from the galactic center, within which the metallicity is close enough to the solar one. However, the relation between the metallicity and the probability to form habitable planets is not quantitatively known and these precise limits should not be considered as robust estimates at all. Moreover, the presence of a giant planet inside the HZ allows the existence of habitable satellites or Trojans³, and the existence of a hot Jupiter (i.e., a giant planet with a very short period) may still allow the formation of habitable planets in the HZ (Raymond et al. 2005).

Since 1995, more than 180 giant extrasolar planets have been discovered (see Chap. 6). All these systems are extremely different from the solar system as they host at least one short period giant planet, with a high eccentricity in most of the cases. This may indicate formation processes that differ from what occurred in the protosolar nebula. However, present techniques favor the detection of short period planets and do not allow yet the detection of planets at the orbital distance of Jupiter. The discovery of Jupiter-like planets, with orbital periods of several years or tens of years is for the near future. From the present surveys and exoplanet detections, we can nevertheless draw the following preliminary conclusions:

- One percent of the (monitored) stars have a detectable planet with a period of less than 10 days (hot Jupiters). The formation of these planets is not fully understood yet. It probably implies the migration of the planet from the outer protoplanetary disk. Such migration would occur through the HZ. According to the authors, this may or may not prevent the formation of habitable planets (see Raymond et al. 2005).
- Seven percent of the stars have a detectable planet with an orbital period shorter than 10 years (this percentage represents a lower limit and increases with the sensitivity of the detection methods and the integrated time of observations).

³ A Trojan is an object orbiting in the Lagrange point (L4 or L5) of another larger object. A habitable Earth-sized planet could orbit in the Lagrange point of one of the giant planets found in the HZ of its star.

- Eccentricities are large in most of the cases, except for the hot Jupiters that have circular orbits.

When observational techniques will be able to detect planets similar to Jupiter, we will then learn more about the frequency of systems like the solar system and their habitability.

Some of the discovered giant planets have quasicircular orbits inside the HZ of their star (see Fig. 7.5). If we cannot consider gaseous giant planets themselves as habitable for life as we know it, we can speculate about the existence of habitable satellites that may orbit around them. Indeed, in the solar system, all the giant planets have satellites and the most massive planets have the largest satellites (Ganymede/Jupiter, Titan/Saturn). Let us consider the system HD 28185, located at about 40 pc from the Sun. The star is very similar to the Sun and hosts a planet with a minimum mass of $5.7 M_{\text{Jup}}$, an orbital distance of about 1 AU, and a small eccentricity ($e = 0.06$). If this planet is surrounded by

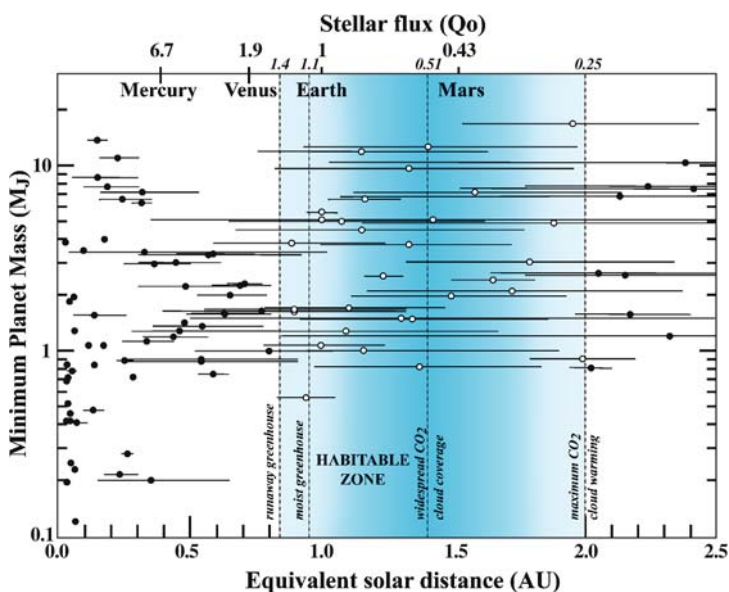


Fig. 7.5. Giant extrasolar planets and habitable zones. This figure gives the minimum mass and the equivalent solar distance for 102 of the known exoplanets. These planets have been detected by the radial velocity technique that allows us to infer their period, eccentricity and minimum mass. *Black lines* represent variations of orbital distance, due to the eccentricity. Taking into account the luminosity of the parent star allows to scale the orbital distance into an equivalent solar distance and thus to compare their position relative to the habitable zone. We can see that some of the detected planets are permanently or periodically inside the HZ of their star. Although, these giant planets are not themselves habitable for life as we know it, big terrestrial satellites able to sustain life may orbit them. Figure by Williams, adapted from Williams and Pollard (2002)

large satellites, those might be habitable (Williams et al. 1997). The detection of habitable satellites is however far from being done with current observation techniques, except maybe for transiting giant planets (Sartoretti and Schneider 1999) for which the precise timing of the transits could allow us to infer the weak perturbation due to satellites.

7.4 The Influence of the Giant Planets on the Habitability of the Terrestrial Planets

Once Jupiter was formed, during the first ten million years of the solar system, it had a major influence on the formation and evolution of the other planets. First of all, by perturbing the orbits of the planetary embryos, it accelerated the formation of the Earth, which reached its present mass in less than 50Ma according to both geochemical data and numerical formation models. In the absence of Jupiter, the same models give a formation time closer to 200Ma and, according to some authors, they produce smaller (Mars-sized) and more numerous terrestrial planets (Raymond et al. 2004; Levison et al. 2003). As discussed earlier, habitability obviously depends on the mass of a planet, due to the need to prevent the loss of atmospheric gases to space and to maintain volcanic activity and CO₂ recycling during a significant fraction of the planet's lifetime. The small size of Mars and the absence of planet between 1.5AU and Jupiter's orbit is probably a consequence of the gravitational influence of Jupiter (Lunine et al. 2003).

The giant planets may also influence the water content of terrestrial planets. If the Earth had accreted only some material formed at 1AU from the Sun, it would be considerably dryer than it is. Thus, some water-rich material that condensed farther away in the solar system (beyond 2.6AU for hydrated minerals and beyond about 4AU for ice) was involved in the formation of our planet. The scenario that emerged recently is that water was brought very early in the history of the solar system, probably by one or a few of the 10 to 100 planetary embryos that collided to form the Earth. These embryos are formed very fast (in less than 1Ma) from material within a very thin range of orbital distance. Their composition is determined by their formation region. According to this scenario, at least one embryo formed beyond 2.6AU was accreted by the proto-Earth. Simulations show that the final water abundance of a terrestrial planet does not simply depend on the orbital distance of the planet: as the water is brought by a small number of outer embryos, ending with enough water to provide habitable conditions is also a question of luck. Raymond et al. (2004, 2005) have simulated the formation of terrestrial planets, and estimated their final water content, in systems with various types of giant planets (including hot Jupiters). If some of the terrestrial planets formed dry in these simulations, most of them contain even more water than Earth. To perform these numerical simulations, a distribution of the planetary embryos is assumed, as well as a relation between their orbital

distance and their water content. These assumptions are reasonable but weakly constrained by observations. The results should thus be considered with caution but they interestingly question the habitability in a new and unexpected way: is too much water a problem? Planets containing 10 or 100 times more water than Earth, “ocean planets” or “water worlds”, would obviously have different tectonic and volcanic activity. They might have no emerged land and no carbonate-silicate cycle, on which relies our present definition of the habitable zone. An ocean-covered planet may not be a place that favors the prebiotic chemistry at the origin of life. Indeed, there would be no way to concentrate organics in a “prebiotic soup” and in some cases, an ice layer would lie underneath the liquid ocean (Léger et al. 2004) preventing the formation of a hydrothermal system.

The properties of the giant planets have also an effect on the bombardment of the terrestrial planets by asteroids and comets. A recent study (Gomes et al. 2005) suggests that the late heavy bombardment (LHB) experienced by the terrestrial planets 4.0 to 3.8Ga ago was triggered by a resonance between Jupiter and Saturn. According to this model, Jupiter and Saturn formed on a circular orbit and have remained on such configuration during the first 800Ma of the solar system. Due to perturbations by planetesimals from the outer solar system, Saturn eventually crossed until it reached a 2:1 resonance with Jupiter. This event perturbed the whole solar system and resulted in the present eccentricity and inclination of Jupiter and Saturn (Tsiganis et al. 2005), and in a peak of bombardment 3.8Ga ago. After the accretion of the Earth involving giant impacts with embryos that ended 4.4Ga and before this dramatic event, the rate of impact on Earth could have been as low as today, or even lower. Was life present on Earth before the LHB? Did the LHB contribute to the emergence of life on Earth by providing extraterrestrial organics or, on the contrary, did it frustrate its appearance and evolution? We do not have the answers to these questions but we can predict that the history of the bombardment of terrestrial exoplanets will be significantly different, due to different properties and orbital configurations of the giant planets (Levison and Agnor 2003), and will affect their habitability in various ways.

7.5 Discussion

The presence of a planet inside the HZ of its star does not imply that this planet is habitable. Mars’ orbit is inside the present HZ of the Sun but would require a much denser atmosphere to sustain permanent liquid water at its surface. Photosynthetic activity requires surface liquid water, surface liquid water requires being inside the HZ. But being inside the HZ does not imply surface liquid water and surface liquid water does not imply life.

However, according to what we know about the Earth, a planet inside the HZ of a Sun-like star, with a mass and composition similar to Earth’s one, should sustain similar tectonic process, a dense atmosphere and a liquid ocean.

Extrapolation to other planetary masses, chemical composition or stellar type remains extremely speculative.

The habitable zone definition we chose applies to planetary surface and thus relies on orbital distance, stellar luminosity and atmospheric properties. Current questions about the possibilities of life in an ocean inside Europa (satellite of Jupiter) or in the Martian subsurface show well that the concept of HZ can be extended. Non-radiative sources of energy can sustain subsurface liquid water. In the case of Europa, the dissipation of tidal forces due to the eccentric orbit of the satellite around Jupiter accounts for the presence of an internal ocean. Following this example, we can define around giant planets a HZ based on tidal heating (Reynolds et al. 1987).

The concept of habitability can also be revisited and enlarged in the light of the recent discoveries of ecosystems adapted to *extreme* environments. Lithotrophic bacteria, for instance, do not need light and can grow inside rocks provided that interstitial water is present. In this case, habitability relies on the geothermal heat flow. Ejected from its orbit and from the HZ of the Sun, the Earth would still host life in its crust. However, we still have no precise idea about the required conditions for the emergence of life. This remains a major uncertainty on habitability: a planet should first satisfy the conditions for life to appear, and those may be weakly related with the definition of the HZ.

7.6 Conclusion and Perspectives

From the viewpoint of the astronomer, the concept of habitability has to be considered with regards to available observation techniques. In our solar system, observation allowed us to recognize the absence of liquid water at the surface of all the planets but the Earth. It also gave us enough physical and chemical indications to infer the structure of some of the planets and satellites and thus to discuss their internal habitability. The presence of liquid water is now strongly suspected in the subsurface of Mars and below the icy crust of Europa, which makes these two bodies high-priority targets for future astrobiological missions.

Yet, observation provides us very few data to evaluate the habitability in other planetary systems, as the detection of terrestrial exoplanets will have to wait for a next generation of instruments. Corot (CNES, Rouan et al. 1998) and Kepler (NASA, Borucki et al. 1997) should detect terrestrial planets around distant stars thanks to their periodic transit in front of their parent star. These missions will give us statistical information about the abundance of terrestrial planets but we will have to wait for Darwin and TPF-C to be able to see planets around nearby stars and to characterize them with low-resolution spectra. These space telescopes should reveal the presence of H₂O, CO₂, and will allow us in some cases to infer the radius and the temperature of the planets and thus to evaluate their potential for life. If life in the universe is abundant enough, we may also detect a biological activity when the chemical composition of the

atmosphere is maintained far from a photochemical equilibrium by an active biosphere (for instance the simultaneous presence of O_2 and CH_4 would be a robust biosignature).

Until these promising missions, radial velocity surveys will give us precious information about the abundance of Jupiter-like planets. This will be the starting point to unveil this fascinating question: is the solar system, which contains at least one habitable planet, a common system in the universe or, conversely, a rare or even unique feature?

Acknowledgements

The author wishes to thank James F. Kasting and Darren M. Williams for providing some of the illustrations for this chapter.

References

- Baraffe, I., Chabrier, G., Allard, F., Hauschildt, P. H., 1998, Evolutionary models for solar metallicity low-mass stars: mass-magnitude relationships and color-magnitude diagrams, *Astron. Astrophys.* 337, 403–412
- Beichman, C. A., Woolf, N. J., Lindensmith, C. A. (eds.), 1999, *The Terrestrial Planet Finder (TPF): a NASA Origins program to search for habitable planets*, JPL Publications, Pasadena
- Borucki, W. J., Koch, D. G., Dunham, E. W., Jenkins, J. M., 1997, The Kepler mission: A mission to determine the frequency of inner planets near the habitable zone of a wide range of stars, In: *Planets beyond the solar system and the next generation of space missions*, p. 153, ASP Conf. Ser. 119
- Catling, D. C., Glein, C. R., Zahnle, K. J., McKay, C. P., 2005, Why O_2 is required by complex life on habitable planets and the concept of planetary oxygenation time, *Astrobiology* 5–3, 415–438
- Cockell, C. S., 1999, Life on Venus, *Plan. Space Sci.* 47, 1487–1501
- Dvorak, R., Pilat-Lohinger, E., Schwarz, R., Freistetter, F., 2004 Extrasolar Trojan planets close to habitable zones, *Astron. Astrophys.* 426, L37–L40
- Favata, F., Roxburgh, I. W., Galadi, D. (eds.) 2001, *Proceedings of the First Eddington Workshop on Stellar Structure and Habitable Planet Finding*, ESA SP-485, ESA Publications Division, Noordwijk
- Forget, F., Pierrehumbert, R. T., 1997, Warming early Mars with carbon dioxide clouds that scatter infrared radiation, *Science* 278, 1273
- Forget, F., 1998, Climate and habitability of terrestrial planets around other stars, dans ERCA vol. 3: From urban air pollution to extra-solar planets, EDP Sciences, 393–407
- Franck, S., Von Bloh, W., Bounama, C., Steffen, M., Schönberner, D., Schellnhuber, H.-J., 2000, Determination of habitable zones in extrasolar planetary systems: Where are Gaia's sisters ?, *JGR* 105-E1, 1651–1658
- Gladman, B. J., Burns, J.A., Duncan, M., Lee, P., Levison, H. F., 1996, The exchange of impact ejecta between terrestrial planets, *Science* 271, 1387

- Gomes, R., Levison, H. F., Tsiganis, K., Morbidelli, A., 2005, Origin of the cataclysmic late heavy bombardment period of the terrestrial planets, *Nature* 435, 466–469
- Gough, D. O., 1981, Solar interior structure and luminosity variations, *Sol. Phys.* 74, 21
- Guinan, E., Ribas, I., 2002, Our changing sun: The role of solar nuclear evolution and magnetic activity on Earth's atmosphere and climate, *ASP Conf. Ser.* 269
- Hart, M. H., 1979, Habitable zones around main sequence stars, *Icarus* 37, 351–357
- Hessler, A. M., Lowe, D. R., Jones, R. L., Bird, D. K., 2004, A lower limit for atmospheric carbon dioxide levels 3.2 billion years ago, *Nature* 428, 736–738
- Holm, N. G., Andersson, E. M. 1998, In: *The Molecular Origins of Life*, Cambridge University Press, Cambridge, p. 86–99
- Horneck, G., Bucker, H., Reitz, G., 1994, Long-term survival of bacterial spores in space, *Adv. Space Res.* 14, 41–45
- Joshi, M. M., 2003, Climate model studies of synchronously rotating planets, *Astrobiology* 3–2, 415
- Joshi, M. M., Haberle, R. M., Reynolds, R. T., 1997, Simulations of the atmospheres of synchronously rotating terrestrial planets orbiting M dwarfs: conditions for atmospheric collapse and the implications for habitability, *Icarus* 129–2, 450–465
- Kasting, J. F., 1988, Runaway and moist greenhouse atmospheres and the evolution of Earth and Venus, *Icarus* 74, 472
- Kasting, J. F., Whitmire, D. P., Reynolds, R. T., 1993, Habitable zones around main sequence, *Icarus* 101, 108
- Léger, A., Mariotti, J. M., Mennesson, B., Ollivier, M., Puget, J. L., Rouan, D., Schneider, J., 1996, Could we search for primitive life on extrasolar planets in the near future? The DARWIN Project, *Icarus* 123, 249
- Levison, H. F., Agnor, C., 2003, The role of giant planets in terrestrial planet formation, *Astron. J.* 125–5, 2692–2713
- Lineweaver, C. H., Fenner, Y., Gibson, B. K., 2004, The Galactic Habitable Zone and the age distribution of complex life in the Milky Way, *Science* 303, 56–62
- Lunine, J. I., 1999, *Earth: Evolution of a Habitable World*, Cambridge University Press, Cambridge
- Lunine, J. I., 2001, The occurrence of jovian planets and the habitability of planetary systems, *Proc. Natl. Acad. Sci.* 98, 809–814
- Lunine, J. L., Chambers, J., Morbidelli, A., Leshin, L. A., 2003, The origin of water on Mars, *Icarus* 165–1, 1–8
- MacKay, C. P., Toon, O. B., Kasting, J. F., 1991, Making Mars habitable, *Nature* 352, 489
- Mastrapa, R. M. E., Glanzberg, H., Head, J. N., Melosh, H. J., Nicholson, W. L., 2001, Survival of bacteria exposed to extreme accelerations for panspermia, *Earth Planet. Sci. Lett.* 189 (1/2), 1–8
- Mischna, M. A., Kasting, J. F., Pavlov, A., Freedman, R., 2000, Influence of carbon dioxide clouds on early martian climate, *Icarus* 145, 546–554
- Napier, W. M., 2004, A mechanism for interstellar panspermia, *MNRAS* 348–1, 46–51
- Pavlov, A. A., Kasting, J. F., Brown, L. L., Rages, K. A., Freedman, R., 2000, Greenhouse warming by CH₄ in the atmosphere of early Earth, *J. Geophys. Res.* 105, 11981
- Rasool, S. I., De Bergh, C., 1970, The runaway greenhouse and accumulation of CO₂ in the Venus atmosphere, *Nature* 226, 1037

- Raymond, S. N., Quinn, T., Lunine, J. I., 2004, Making other Earths: dynamical simulations of terrestrial planet formation and water delivery, *Icarus* 168-1, 1–17
- Raymond, S. N., Quinn, T., Lunine, J. I., 2005, The formation and habitability of terrestrial planets in the presence of close-in giant planets, *Icarus* 177, 256–263
- Reynolds, R. T., MacKay, C. P., Kasting, J. F., 1987, Europa, tidally heated oceans, and habitable zones around giant planets, *Adv. Space Res.* 7, 125
- Rosenqvist, J., Chassefiere, E., 1995, Inorganic chemistry of O₂ in a dense primitive atmosphere, *Planet. Space Sci.* 43, 3
- Rouan, D., Baglin, A., Copet, E., Schneider, J., Barge, P., Deleuil, M., Vuillemin, A., Léger, A., 1998, The exosolar planets program of the COROT satellite, *Earth Moon Planets* 81, 79
- Rye, R., Kuo, P. H., Holland, H., 1995, Atmospheric carbon dioxide concentrations before 2.2 billions years ago, *Nature* 378, 603
- Sartoretti, P., Schneider J., 1999, On the detection of satellites of extrasolar planets with the method of transits, *Astron. Astrophys. Suppl.* 134, 553–560
- Schneider, J., 2002, L'Encyclopédie des Planètes Extrasolaires. <http://www.obspm.fr/encycl/f-encycl.html>. Cited 2002
- Selsis, F., 2000, Evolution physicochimique des atmosphères de planètes telluriques. Dissertation, Université de Bordeaux
- Selsis, F., 2002, Occurrence and detectability of O₂-rich atmosphere in circumstellar habitable zones, *ASP Conf. Ser.* 269
- Selsis, F., Léger, A., Ollivier, M., 2003, Signatures spectroscopiques de vie sur les exoplanètes. Les missions Darwin et TPF, dans L'origine de la vie sur Terre et dans l'Univers Tome 2, Presses Universitaires de Bordeaux, 343–369
- Sleep, N. H., 2005, Cutting anthropic knots and the rise of O₂, *Astrobiology* 5-3, 331–332
- Sleep, N. H., Zahnle, K., 2001, Carbon dioxide cycling and implications for climate on ancient Earth, *J. Geophys. Res.* 106, 1373
- Tsiganis, K., Gomes, R., Morbidelli, A., Levison, H. F. (2005) Origin of the orbital architecture of the giant planets of the solar system, *Nature* 435, 459–461
- Volonte, S., Laurant, R., Whitcomb, G., Karlsson, A., Fridlund, M., Ollivier, M., Gondoin, P., Guideroni, B., Granato, G. L., Amils, R., Smith, M., 2000, Darwin: the infrared space interferometer, Technical Report, ESA
- Walker, J. C. G., Hays, P. B., Kasting, J. F., 1981, A negative feedback mechanism for the long-term stabilization of Earth's surface temperature, *J. Geophys. Res.* 86, 9776
- Williams, D. M., Kasting, J. F., Wade, R. A., 1997, Habitable moons around extrasolar giant planets, *Nature* 385, 234–236
- Williams, D. M., Pollard, D., 2002, Earth-like planets on eccentric orbits: Excursions beyond the Habitable zone, *Inter. J. Astrobio* 1(1), 61–69

8 Habitability: the Point of View of a Biologist

Purificación López-García

8.1 Introduction

Defining habitability from a biological point of view is not an easy task. Biologists consider the concept of habitability as intimately linked to the concept of life, which is in itself a subject of debate, and to the limits of life as we know it. Furthermore, if some systems can be classified as unequivocally *living*, other entities occupy an ambiguous and undetermined position. In fact, the term “habitability” is not frequently used in biology. Biologists rely on an experimental observation, *life exists*, and make of it their object of study. Consequently, biologists talk about habitats and ecological niches rather than habitability. Nevertheless, the discovery of microbial life in extreme environments opened new questions about the type of settings that can be habitable, i.e., colonized by known life forms. These concern not only terrestrial habitats that were thought sterile for a long time, but also planets or satellites that might offer similar conditions to those under which life is able to thrive on Earth.

8.1.1 The Concept of Habitability

The idea of habitability was first introduced by Dole in the 1960s, and referred to the planetary conditions that allowed for human life (Dole 1964). However, in course of time, the term of habitability became less astringent and anthropocentric, pertaining to the stability of liquid water at the surface of a planet, and being mostly used by astronomers and astrobiologists. Habitability turned into a utilitarian concept subject to the constraints imposed by the available methods to make observations at great distances. Thus, a habitable zone was defined as the region around a star where temperature and pressure are such that liquid water can exist on a planet’s surface (Franck et al. 2001; Kasting et al. 1993) (see also the Chap. 7). More recently, the idea of a habitable zone is being enlarged to include also the existence of stable liquid water at a planet’s subsurface, which could eventually sustain a deep subterranean biosphere. In this latter case, the heat sources allowing the occurrence of liquid water will encompass not only the proximity to a star but also internal sources derived from gravitational forces or radioactivity. Bodies such as the Jupiter’s satellite Europa, where a liquid ocean could exist under the ice cover (Carr et al. 1998; Kivelson et al. 2000),

and Mars, which may contain liquid water in the subsurface (Malin and Edgett 2003; Bibring et al. 2004), would then be considered as habitable zones.

8.1.2 Habitability in Biological Terms

In biology, the term habitability is not actually used. Nonetheless, one can try to define what habitability should be for a biologist in a very instinctive way. Habitability would refer to the set of necessary and also sufficient conditions for life to be able to exist in an active form (not in a dormant state). In principle, a non-inhabited, habitable environment could become inhabited if it was colonized by an external life form (e.g., by panspermia). The question that arises here is whether habitable planets become necessarily inhabited or not, or in other words, whether the role of necessity prevails over that of chance. At any rate, determining habitability conditions involves defining life. Only then, will we be able to (1) identify living beings, (2) establish their common properties, and (3) infer which essential minimal conditions are required for their existence.

8.2 What is Life?

8.2.1 Life's Definitions

There are several definitions of life, but none is completely satisfying. Whatever the criterion used to build a particular, arbitrary, life definition, one can always find systems in nature for which the classification living/non-living is problematic or uncertain. In general, we are bound to define life by the properties that we ascribe to it, that is, to a complex system that is far from equilibrium and that follows the laws of thermodynamics (Elitzur 1994). In this context, two main definition types can be distinguished:

1. The first definition type highlights the properties of autoreplication and evolution, and corresponds to the definitions officially adopted by NASA: life is a self-sustaining chemical system capable of undergoing Darwinian evolution (Joyce 1994). *Darwinian* implies that the offspring may have modifications (derived from mutations in the genetic material, i.e., from a molecular evolution) upon which the laws of natural selection may act. In this way, some individuals with specific traits may gain an advantage over others and hence be more successful in producing offspring.
2. The other definition type is based on autocatalysis or *autopoiesis* (Luisi 2003; Varela et al. 1974). In this model, the idea of vesicles, confined structures, bounded by an envelope and capable of metabolism is essential: the minimal life form is a self-organizing and self-sustaining system bounded by a semi-permeable envelope that, owing to an inner network of reactions, re-generate all the system's components.

These definitions are more or less useful, but they are not adapted to all cases. For instance, under a definition based on the properties of auto-replication and evolution, certain crystals able to incorporate and replicate defects in their growth could be considered as living. In this sense, Cairns-Smith proposed that life evolved initially from clay minerals, which had the capability to grow following an ordered pattern and to incorporate and duplicate imperfections (Cairns-Smith 1977). Could some crystals be considered as living? On the opposite side, some animals derived from crossings between close species, such as mules (derived from a donkey and a horse), cannot reproduce and would not fit the above definition of life.

8.2.2 Is it Living?

The two most ambiguous and problematic examples regarding those life definitions are probably viruses and what is called “artificial life”.

Viruses are obligatory parasites of cells. They are composed of two major constituents: (1) a genome made of nucleic acid (it can be RNA or DNA, simple or double strand), and (2) one capsid made of proteins, sometimes also enveloped by lipids. Viruses exhibit an extraordinary diversity (Morse 1993; Murphy 1995). Some RNA viruses have very small genomes (0.22kbp)¹ (Collins et al. 1998) whereas, on the contrary, some double-stranded DNA viruses have genomes as large as 800kbp, which is comparable in size to that of genomes of small parasitic bacteria (La Scola et al. 2003). However, viruses cannot make their own proteins, as they do not contain ribosomes, and similarly, they are unable to duplicate their own genetic material. Although they usually carry the genes encoding the polymerase to do it, these polymerases must be synthesized by cells. Therefore, viruses could not exist without cells, because they are not autonomous and cannot replicate themselves. Furthermore, they are unable to gain energy, since they lack any form of energy metabolism. Consequently, viruses do not fit any of the given definitions of life. In this sense, they could be considered as entities placed at the frontiers between living and non-living. Nevertheless, viruses are obvious indicators of the existence of cellular life. Of course, this type of argument can be also applied to any obligatory parasite in general, which shows the ambiguity of existing life definitions. Nevertheless, although cellular parasites have lost many functions they no longer need, they retain the basic machineries allowing both, the replication of their own genetic material and their protein synthesis. These machineries and functions do not exist in viruses.

An opposite, but also problematic case that exemplifies the ambiguity of current life definitions is that of artificial life, which refers to the software programs

¹ Kilobase pair (kbp), a unit of double-stranded nucleic acid length measurement; 1 kbp is equivalent to 1000 nucleotides in a DNA or RNA duplex. For single-stranded molecules, nucleotides or bases rather than base pairs are used as length measurement units.

that mimic life processes (see Chap. 15). They transfer and modify information, in an analogous way to biological life. In biological systems, information can be transferred within a single cell (from DNA to RNA and from RNA to proteins), or from one cell to its offspring (vertical inheritance), or else, from one cell to their co-existing mates (horizontal gene transfer). Should we then consider *artificial life* as *true life*? It may be puzzling to note that if (biological) viruses are unambiguous indicators of (cellular) life, informatic viruses could in an analogous manner attest for the existence of such artificial life.

8.3 The Cell

Besides those problematic cases, the biologist may choose to use a conservative and empirical criterion to define the properties that are inherent to life as we know it by considering only cellular organisms. In this sense, the cell becomes the fundamental unit of life. From a thermodynamic point of view, a cell is an open system that exchanges matter and energy with the outside.

8.3.1 Properties

Cells are complex chemical systems possessing the following basic properties (Fig. 8.1):

- Energy *metabolism* and nutrition. Cells incorporate chemical compounds from the environment, transform them and eliminate other chemical compounds as residues. In this way, cells obtain the rough material to synthesize

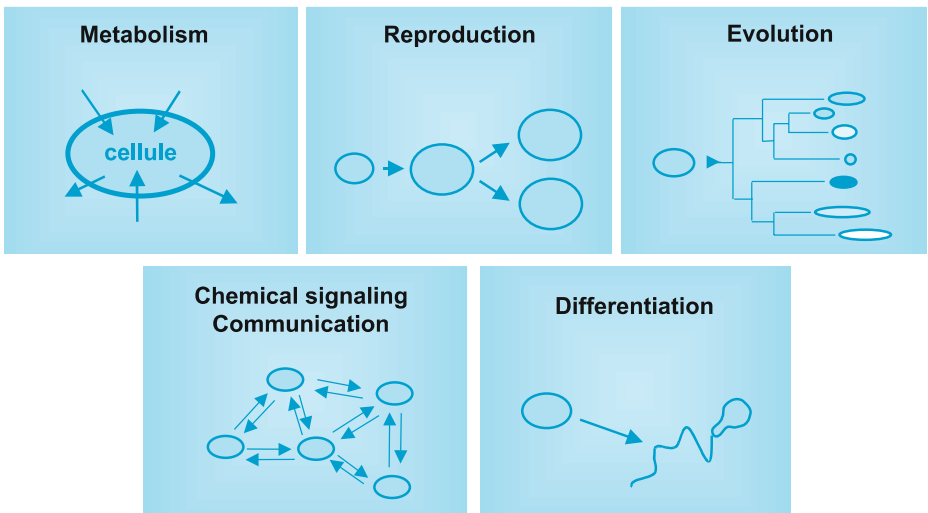


Fig. 8.1. Major properties of the cell

their own cellular components, and gain the energy needed to carry out all biochemical reactions.

- *Reproduction* (growth and replication). Cells duplicate their genetic material and divide to produce two cells similar to the mother cell.
- *Evolution*. Cells undergo mutations at the level of their genetic material that lead, as a consequence, to the modification of their offspring upon which natural selection may act.
- *Communication*. Cells can respond to physical and chemical environmental stimuli and can communicate among them through small signaling molecules.
- *Differentiation*. Some cells can modify their structure during their life cycle. Examples of this are the formation of spores, gametes, and the differentiation to form a particular tissue.

8.3.2 Prokaryotes and Eukaryotes

On Earth, two types of cellular structures are known: the prokaryotic cell and the eukaryotic cell (Fig. 8.2). Prokaryotes have generally a simpler organization, with the genetic material (one or more DNA chromosomes) directly immersed in the cytoplasm, containing also the ribosomes. The prokaryotic cell is bounded by a double membrane, the plasma membrane and, very often, a cell wall. Some prokaryotes possess an additional external membrane, others have an external proteinaceous layer, the S layer.

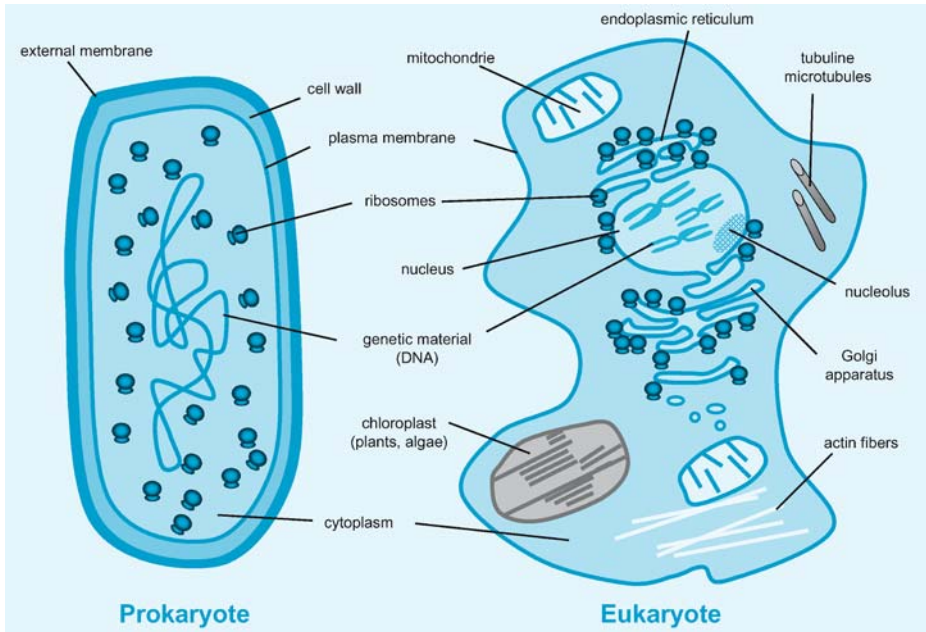


Fig. 8.2. Schematic view of the cell structure in prokaryotes and eukaryotes

The most important structural differences exhibited by eukaryotes in comparison with prokaryotes are:

1. The presence of a membrane-bounded *nucleus* that separates the genetic material from the cytoplasm. The nucleus membrane system can extend to form the endoplasmic reticulum, upon which are active ribosomes.
2. The presence of an internal *cytoskeleton* composed mostly of actin fibers and tubulin microtubules. Its role is the maintenance of the cellular organization and morphology. It also facilitates the traffic of internal components.
3. The presence of membrane-bounded organelles, notably *mitochondria* and, in the case of photosynthetic eukaryotes (algae, plants), *chloroplasts*. There is now compelling evidence that these organelles derive from ancient endosymbiotic bacteria. Chloroplasts derive from cyanobacteria, and mitochondria derive from proteobacteria of the alpha subdivision (Gray 1989).

8.3.3 The Tree of Life

The two types of cellular architecture, prokaryotic and eukaryotic, do not perfectly superimpose the phylogenetic (or genealogical), natural classification of organisms. The prokaryotic world is divided into two large groups that are as genetically distant (and therefore evolutionary distant) from one another as each of them is from eukaryotes (Fig. 8.3). All cellular life known on Earth is divided into three major domains, Archaea, Bacteria, the two prokaryotic domains, and Eucarya (the eukaryotes). In spite of a similar morphology and mode of life, archaea and bacteria have important differences in their molecular biology. In fact, archaea and eukaryotes are quite similar in their mechanisms to replicate DNA, to transcribe DNA to RNA, and to translate proteins (Keeling et al. 1994). Eukaryotes display the largest morphological diversity (plants, animals, fungi and many lineages of eukaryotic microorganisms such as ciliates, dinoflagellates, algae, foraminifers, radiolarians, etc.). In turn, prokaryotes display the largest physiological and metabolic diversity (see below).

The evolutionary origin of eukaryotes is a subject of active controversy. Most scientists sustain the idea that eukaryotes, more complex, derive from simpler prokaryotes via a complexification process. However, a few authors think the other way round: prokaryotes would derive from ancestral eukaryotes via a reduction process (Forterre 1995; Reaney 1974). The existing models to explain the origin of eukaryotes from prokaryotic ancestors are varied, and hotly debated. Nevertheless, two major currents can be distinguished. One proposes that present-day eukaryotes originated from a third ancestral lineage of prokaryotes that was phylogenetically closer to that of the archaea. This would explain why archaea and eukaryotes share many similarities. This third lineage would have developed a nucleus and acquired mitochondria by endosymbiosis (Brown and Doolittle 1997; Doolittle 1998). The other current corresponds to that of chimerical models, which propose that eukaryotes originated from symbioses between

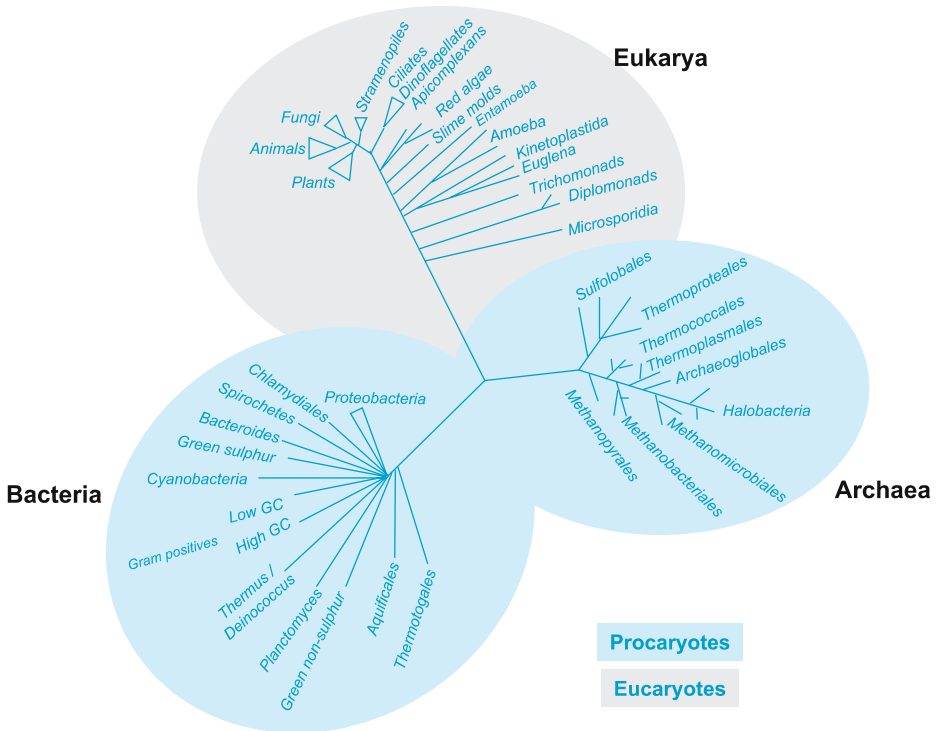


Fig. 8.3. Schematic vision of the universal tree of life based upon 16S/18S ribosomal RNA gene sequences. In a phylogenetic tree, the length of branches leading to different species or lineages of organisms is proportional to the evolutionary distance

archaea and bacteria and, therefore, reject the existence of an independent pro-toeukaryotic lineage (Martin and Müller 1998; Moreira and López-García 1998; Searcy et al. 1978). These hypotheses are very diverse, but all of them are somehow reinforced by the evolutionary fact that eukaryotic organelles do derive from symbiotic bacteria. There is no doubt that symbiosis has played an essential role in the emergence of contemporary eukaryotes.

8.4 Common Denominators of Life on Earth

The strategy that a biologist may use to establish habitability criteria is comparable to that needed to infer which conditions and elements led to the emergence of life on Earth. Starting from life as we know it, we can extract the common features and the basic materials that are essential to any organism, as well as determine the physicochemical limits permissive for life forms analogous to terrestrial organisms. Of course, this is an empirical principle, based on what we consider living, which is also quite conservative. Once these mini-

mal physicochemical requirements for terrestrial life are established, one could argue that the existence of similar conditions in other planets would convert them to habitable, and even prone to the emergence of life. This does not necessarily imply that if those conditions exist in a planet life has developed. On the other hand, one could also imagine that life forms that we would be unable to recognize might exist somewhere in the universe under very different conditions.

Analysis of terrestrial cellular life, (prokaryotes and eukaryotes), provides a basis to try to find common denominators at different scales:

- At the molecular level, i.e., the level of molecules and elements integrating the cell.
- At the cellular level, i.e., at the level of physiology and metabolism, implying the chemical reactions required for cell housekeeping.
- At the ecological level, at the level of cell interrelationships with other cells and with their environment, which would allow to determine the type of habitats where life can develop.

8.4.1 Elements and Molecules

8.4.1.1 Biological Macromolecules

All prokaryotes and eukaryotes, i.e., all cells, possess several elements and macromolecules in common (Madigan et al. 2002) (Fig. 8.4):

- **Proteins**, several of which are conserved in the three domains of life, notably membrane ATPases, ribosomal proteins and those involved in transcription and translation.
- **Lipid** membranes that form semi-permeable barriers to different ions and molecules. Membrane phospholipids are generally disposed in bilayers with the hydrophilic heads (glycerol-phosphate moieties) towards the external surface, and the hydrophobic tails (isoprenoids or fatty acids) towards the interior.
- **Genetic material** composed of *DNA*, either isolated in a nucleus (eukaryotes) or forming a nucleoid in the cytoplasm (prokaryotes).
- **Cytoplasm**, a kind of fluid gel that is a saline solution containing many molecules in solution or in suspension, where *water* is the essential solvent. It serves for metabolite transport and is the place where most biochemical reactions take place.
- **Ribosomes**, made out of *RNA* and proteins, where protein synthesis occurs.
- **Polysaccharides**, essential intermediates in many metabolic pathways and, for instance, as part of glycosylated membrane proteins in the external wall.

Biological macromolecules, proteins, nucleic acids (*DNA* and *RNA*), lipids and polysaccharides can, in turn, be decomposed in simpler molecules: amino acids, nitrogen bases, glycerol, fatty acids, isoprenes and simple sugars. And

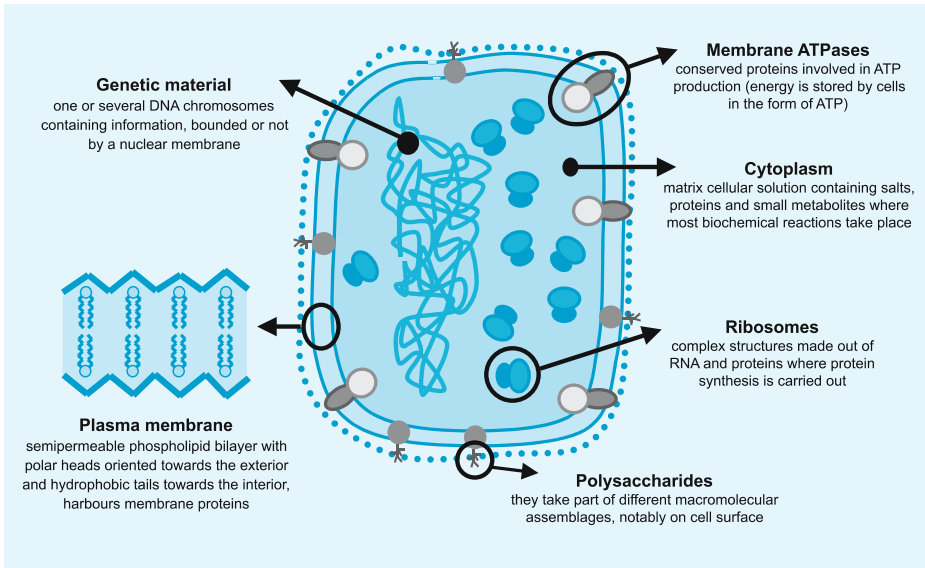


Fig. 8.4. Universal components of cells

again, these are composed by a limited number of elements, basically C, H, O, N, P and S, although other elements (oligoelements) are required for life in low proportions (Fe, Ca, Mn, etc.).

8.4.1.2 A Carbon-Based Chemistry

Carbon is fundamental for biochemistry. Furthermore, carbon chemistry dominates the Universe. Various features convert C in one indispensable element for biochemistry (Lunine 1999):

- A large abundance. Carbon is the fourth most abundant element in the universe after hydrogen, helium and oxygen.
- A high tendency to form covalent links with other elements and with itself, which allows the formation of long chains, rings, etc.
- An important versatility. C forms links very easily with other elements, and also ramifications, which is crucial to form nucleotide and amino acid skeletons.

Silicon resembles carbon in certain properties, and this has raised discussions about the possibility of a life based on Si. Silicon is also very abundant in the universe. It is less abundant than C, but it is the second most important element in the terrestrial crust after oxygen. Si has also a similar capacity to form covalent links (chains and rings). However, it is far less versatile than C (Lunine 1999).

8.4.1.3 Water: the Universal Solvent

Liquid water is indispensable for life. Its role as a solvent in the cytoplasm and other cellular compartments where biochemical reactions and transport of metabolites occurs is primordial. Among the properties that make water an essential molecule are the following:

- Water is very abundant on Earth and on other planets. Oxygen is the third most abundant element in the universe, mostly forming CO (volatile) and water.
- H₂O is liquid within the appropriate temperature limits for most known organic chemistry reactions.
- Water easily forms hydrogen bonds with other molecules, including H₂O molecules.
- Water has a marked bipolar character. H₂O molecules are good electric conductors as well as a solvent for many organic molecules.

In addition to H₂O, are there other small molecules necessary for life? Is O₂ required for life? Although O₂ is used for respiration by a large fraction of contemporary organisms, molecular oxygen is not absolutely needed for life. Some organisms cannot even tolerate its presence (strict anaerobes). Furthermore, O₂ was present only in trace amounts for a long period in Earth's history, during which microbial life dominated, before the atmospheric O₂ content increased to high levels by the action of oxygenic photosynthesizing cyanobacteria about 2 Ga ago.

8.4.2 Cellular Metabolism

A cell needs a *carbon* source to make its own components and an *energy* source to feed biochemical reactions (Fig. 8.5).

The source of carbon can be of two types. Either it is organic and comes from organic carbon already fixed by other cells (an organic carbon source could also be envisaged in the context of prebiotic synthesis at the time when life originated), or it is inorganic. Carbon is then fixed from CO₂ or other C1 compounds. Regarding the energy source, it may be obtained from light or from oxido-reduction (redox) chemical reactions. Cells store energy in ATP molecules, since the nucleotide ATP contains two high-energy phosphate bonds.

Carbon can be transformed into (fixed in) organic matter by photosynthesis (a pathway coupled to the production of energy from light) or by chemosynthesis (via metabolic pathways coupled to the production of energy from redox reactions). Similarly, the processes known to obtain energy for cellular metabolism are the following:

- **Respiration.** Oxido-reduction reaction to produce energy (ATP) in which both, the electron donor, and acceptor may be organic or inorganic molecules.

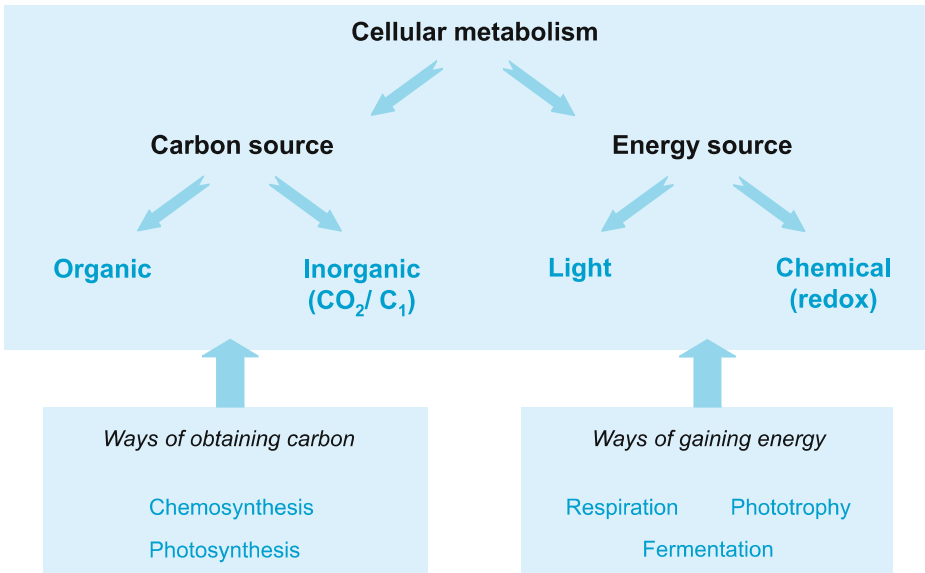


Fig. 8.5. Metabolic pathways to obtain carbon and energy

The respiration of organic molecules with molecular oxygen is widespread, but there are many other respiration types, notably including different final electron acceptors: nitrate, sulfate, FeIII, MnIV, etc. (see below).

- **Fermentation.** Energy-delivering partial oxidation of organic molecules under anaerobic conditions. It is the type of metabolic reaction carried out by bacteria and fungi involved in beer, wine and vinegar fabrication, with alcohols and acids as final products.
- **Phototrophy.** Way of obtaining energy from light. Most phototrophic organisms also fix inorganic carbon (photosynthesizers), but some utilize organic molecules as a carbon source (Table 8.1). Photosynthesis can be oxygenic if it leads to the formation of O₂ as a waste product from water, or anoxygenic, when the final product is not O₂ (normally S and derivatives, but also FeIII).

A classification of organisms as a function of the energy and carbon sources is schematically given in Table 8.1.

Prokaryotes are metabolically versatile and diverse. We can find all metabolic types within prokaryotes. Of course, some specificities exist; for instance, methanogenesis (the energy-yielding production of CH₄ from H₂ and CO₂ or acetate) is typical of some archaea, whereas oxygenic photosynthesis is restricted to cyanobacteria. On the opposite side, eukaryotes can be only either heterotrophs, which is their ancestral state, or photoautotrophs, in the case of those that acquired endosymbiotic cyanobacteria that later evolved into chloroplasts.

Table 8.1. Classification of living organisms depending on their way of gaining energy and carbon

	Carbon source	Energy source	Examples
Photoautotrophs	CO ₂	Light	Cyanobacteria, plants
Photoheterotrophs	Organic	Light	Heliobacteria, some halophilic archaea
Chemo- (litho)autotrophs	CO ₂	Redox reaction using inorganic compounds	Methanogenic archaea, sulfate-reducing bacteria
Chemoheterotrophs	Organic	Oxidation of organics	Animals, fungi, many bacteria

If we analyze the different types of energy metabolism, all energy-yielding reactions can be summarized as redox reactions involving the transport of electrons from a donor to an acceptor. Even if photosynthesis is somewhat particular, it shares at a certain point a similar *modus operandi*, since light energy is transformed in redox chemical energy. Photosynthesis implies the transport (cyclical or not) of electrons, once excited by photons, and the generation of “reduction power” in the form of NAD(P)H. The latter can be oxidized to NAD(P)⁺, serving to reduce CO₂ to make hydrocarbons. The external electron donors can be H₂O, in oxygenic photosynthesis, or H₂S, S₂O₃²⁻, S⁰ or Fe²⁺, in anoxygenic photosynthesis (Madigan et al. 2002). Looking at known chemical redox reactions, the number of possible electron donors and acceptors is amazing (Table 8.2). Organisms appear to gain energy from any reaction with a negative ΔG. Nevertheless, most likely we do not know yet all the possible redox couples that can be exploited by nature. The possibility that simple electrodes (anodes) may serve as electron acceptors has been shown (Bond et al. 2002), and more recently, the same has been observed for graphite electrodes as electron donors (cathodes) (Gregory et al. 2004).

Table 8.2. Electron donors and acceptors used in metabolic redox reactions that are known to date

Electron donors	Electron acceptors
Reduced organic compounds, H ₂ O, H ₂ S, S ₂ O ₃ ²⁻ , S ₀ , H ₂ , Fe ²⁺	Fumarate and other oxidized organic compounds, DMSO, NA(D)P+, O ₂ , S ₀ , NO ₃ ⁻ , SO ₄ ²⁻ , CO ₂ ,
CH ₄ & C ₁ -derived, NH ₄ ⁺ , NO ₂ ⁻ , Mn ²⁺ , FeS ₂ , FeCO ₃ , HPO ₃ ²⁻	Fe ³⁺ , CrO ₄ ⁴⁺ , oxidized humic acids, Mn ⁴⁺ , UO ₂ ²⁺ , SeO ₄ ²⁻ , AsO ₄ ³⁻ , DMSO, trimethylamine

8.4.3 The Limits of Terrestrial Life

The observation and determination of physicochemical, environmental limits within which life is able to develop on Earth constitutes an empirical base to establish minimal habitability criteria for other planets. Historically, biologists have studied macroscopic life, easy to observe, in several different habitats (fresh-water and marine, as well as many terrestrial ecosystems). Microbiologists had also focused their studies in those habitats that were only apparently colonized by living organisms. However, during the last three decades, microbiologists have extended their studies to extreme environments, that is, environments that are characterized by physicochemical conditions that were considered hostile for life. In this way, we realized that the range of physicochemical parameter values allowing life on Earth is much more extended than was previously thought. Many microorganisms specifically adapted to extreme environments, the so-called extremophiles, have been discovered (Table 8.3). Extreme environments are generally characterized by a reduction in species diversity, an overwhelming dominance of microbial life, and a frequently increased role of microbial-mineral interactions (López-García 2005; Rothschild and Mancinelli 2001).

Regarding life under extreme conditions, some parameters appear to be more restrictive than others (Fig. 8.6). Temperature is among the most influential factors. We now know that some organisms display metabolic activity on Earth at nearly -20°C in permafrost or glacier environments (Rivkina et al. 2000) and recent measurements even suggest that some kind of metabolic activity to allow survival could exist down to -40°C (Price and Sowers 2004). On the other hand, some microorganisms can grow at temperatures of up to $+113^{\circ}\text{C}$ and,

Table 8.3. Classification of extremophiles

Physicochemical parameter	Type of organism	Definition/Optimal growth
Temperature	Hyperthermophile	$> 80^{\circ}\text{C}$
	Thermophile	$60\text{--}80^{\circ}\text{C}$
	Psychrophile	$< 10\text{--}15^{\circ}\text{C}$
pH	Acidophile	$\text{pH} < 2\text{--}3$
	Alkaliphile	$\text{pH} > 9\text{--}10$
Salinity	Halophile	High salt concentrations ($\sim 2\text{--}5\text{M NaCl}$)
Pressure	Barophile (Piezophile)	High pressure
Desiccation	Xerophile	Enduring strong dryness
Radiation	Radiotolerant	Tolerate high radiation doses
Metals	Metallotolerant	Tolerate high metal concentrations

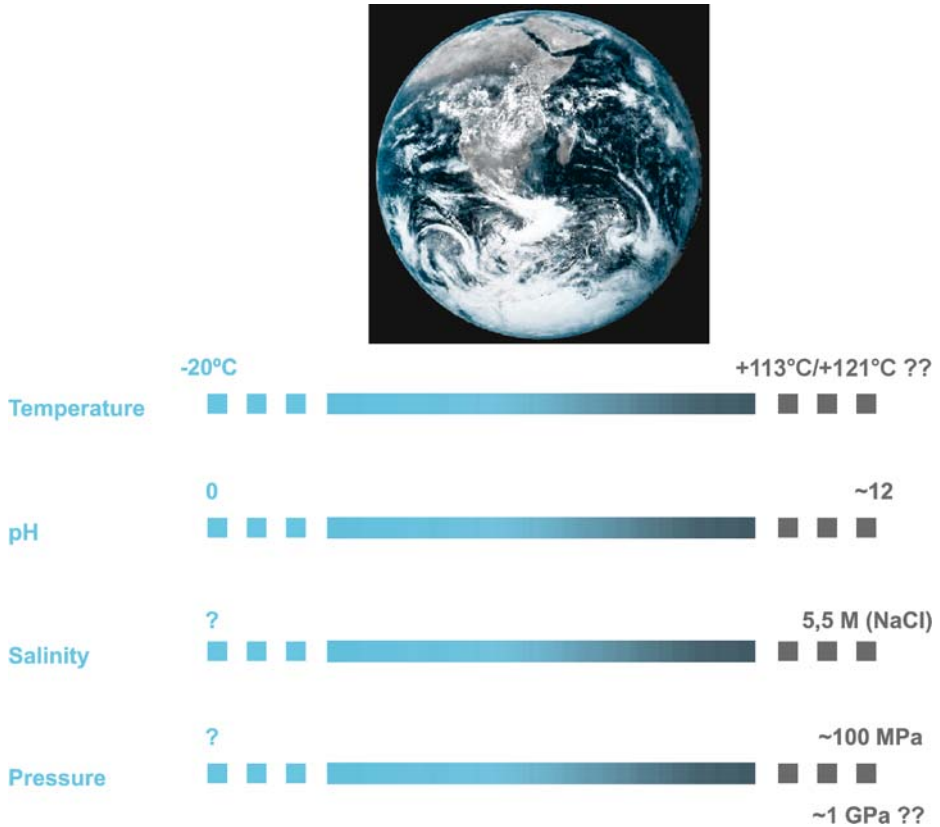


Fig. 8.6. Some known limits for life on Earth

although controversial, even 121 °C (upper growth limit reported for a hyperthermophilic archaeon isolated from the East Pacific Rise; Kashefi and Lovley 2003). It appears that the availability of liquid water is crucial too. Nevertheless, water can be liquid in microniches within ice (e.g., at small fissures and crevices where salt concentrates, or beneath glaciers at high pressure), and also at the bottom of the ocean where pressure is very high. Temperature appears to be also one of the limiting factors for life in the deep subsurface (Furnes et al. 1999; Pedersen 2000). Regarding pH, we know about the existence of microorganisms able to thrive at pH 0 (Schleper et al. 1995) and at high pH in soda lakes, up to values of ~12 (Jones et al. 1998). The fact that we do not know organisms living at even higher pH values might be just related to the absence of naturally occurring environments at pH above 13. High salt concentration is also an enduring condition for life, because it diminishes the actual availability of water. However, many microorganisms can grow at very high salt conditions, including saturating NaCl concentrations (Oren 1994). Finally, even if high pressures may in principle contribute to constrain life conditions, they seem less determining

than other physicochemical variables. Indeed, life appears to adapt well to the deepest oceanic floors, such as the Marianas Trench (~100MPa) (Takami et al. 1997). It has been reported that some microbial cultures can endure pressures as high as 1 GPa under laboratory conditions (Sharma et al. 2002), although the reliability of this information is disputed (Yayanos 2002).

8.5 Perspectives

From a biologist's point of view, it is difficult to determine a set of habitability conditions, because this will depend on the definition of life that is chosen. Using a conservative criterion and starting from life as we know it, we are closer today to determining which are the physicochemical limits permitting life on Earth and its common denominators than we were a few years ago. Nevertheless, we are likely still far from knowing all possible metabolic strategies and all groups of organisms that inhabit our own planet. At any rate, the fact that life exists in extreme terrestrial environments extends considerably the spectrum of planets where life should be looked for or, in other words, the spectrum of habitable planets. Thus, the search for life in Europa under the ice cover, where putative psychrophilic organisms analogous to those living in permafrost could develop is not completely groundless (Chyba and Phillips 2001). Similarly, it would not be impossible that some microorganisms could still be living in the Martian subsurface (or other planet's subsurface) (White et al. 1998). Life as we know is based on a tripartite foundation: a carbon-based biochemistry, liquid water, and a source of free energy. The combination of the three might not be so uncommon in the universe.

References

- Bibring, J.P., Langevin, Y., Poulet, F., Gendrin, A., Gondet, B., Berthe, M., Soufflot, A., Drossart, P., Combes, M., Bellucci, G., Moroz, V., Mangold, N., Schmitt, B., and The Omega Team (2004) Perennial water ice identified in the south polar cap of Mars. *Nature*, **428**: 627–630.
- Bond, D.R., Holmes, D.E., Tender, L.M. and Lovley, D.R. (2002) Electrode-reducing microorganisms that harvest energy from marine sediments. *Science*, **295**: 483–485.
- Brown, J.R. and Doolittle, W.F. (1997) Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.*, **61**: 456–502.
- Cairns-Smith, A.G. (1977) Takeover mechanisms and early biochemical evolution. *Biosystems*, **9**: 105–109.
- Carr, M.H., Belton, M.J., Chapman, C.R., Davies, M.E., Geissler, P., Greenberg, R., McEwen, A.S., Tufts, B.R., Greeley, R., Sullivan, R., Head, J.W., Pappalardo, R.T., Klaasen, K.P., Johnson, T.V., Kaufman, J., Senske, D., Moore, J., Neukum, G., Schubert, G., Burns, J.A., Thomas, P. and Veverka, J. (1998) Evidence for a subsurface ocean on Europa. *Nature*, **391**: 363–365.

- Chyba, C. and Phillips, C. (2001) Possible ecosystems and the search for life on Europa. *Proc. Natl. Acad. Sci. USA*, **98**: 801–804.
- Collins, R.F., Gellatly, D.L., Sehgal, O.P. and Abouhaidar, M.G. (1998) Self-cleaving circular RNA associated with rice yellow mottle virus is the smallest viroid-like RNA. *Virology*, **241**: 269–275.
- Dole, S.H. (1964) *Habitable Planets for Man*. Blaisdell, New York.
- Doolittle, W.F. (1998) You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.*, **14**: 307–311.
- Elitzur, A.C. (1994) Let there be life. Thermodynamic reflections on biogenesis and evolution. *J. Theor. Biol.*, **168**: 429–459.
- Forterre, P. (1995) Thermoreduction, a hypothesis for the origin of prokaryotes. *C. R. Acad. Sci. III*, **318**: 415–422.
- Franck, S., Block, A., von Bloh, W., Bounama, C., Garrido, I. and Schellnhuber, H.J. (2001) Planetary habitability: is Earth commonplace in the Milky Way? *Naturwissenschaften*, **88**: 416–426.
- Furnes, H., Muehlenbachs, K., Tumyr, O., Torsvik, T. and Thorseth, I.H. (1999) Depth of active bio-alteration in the ocean crust: Costa Rica Rift (Hole 504B). *Terra Nova*, **11**: 228–233.
- Gray, M.W. (1989) The evolutionary origins of organelles. *Trends Genet.*, **5**: 294–299.
- Gregory, K.B., Bond, D.R. and Lovley, D.R. (2004) Graphite electrodes as electron donors for anaerobic respiration. *Environ. Microbiol.*, **6**: 596–604.
- Jones, B.E., Grant, W.D., Duckworth, A.W. and Owenson, G.G. (1998) Microbial diversity of soda lakes. *Extremophiles*, **2**: 191–200.
- Joyce, G.F. (1994) In: Deamer D. W. and Fleischaker G. R. (eds.) *Forward in Origins of Life: The Centary Concepts*. Jones and Bartlett, Boston.
- Kashefi, K. and Lovley, D.R. (2003) Extending the upper temperature limit for life. *Science*, **301**: 934.
- Kasting, J.F., Whitmire, D.P. and Reynolds, R.T. (1993) Habitable zones around main sequence stars. *Icarus*, **101**: 108–128.
- Keeling, P.J., Charlebois, R.L. and Doolittle, W.F. (1994) Archaeobacterial genomes: eubacterial form and eukaryotic content. *Curr. Opin. Genet. Dev.*, **4**: 816–822.
- Kivelson, M.G., Khurana, K.K., Russell, C.T., Volwerk, M., Walker, R.J. and Zimmer, C. (2000) Galileo magnetometer measurements: a stronger case for a subsurface ocean at Europa. *Science*, **289**: 1340–1343.
- La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J.M. and Raoult, D. (2003) A giant virus in amoebae. *Science*, **299**: 2033.
- López-García, P. (2005) Extremophiles. In: Gargaud, M., Barbier, B., Martin, H. and Reisse, J. (eds.) *Lectures in Astrobiology*, vol. I, pp. 657–679. Springer, Berlin Heidelberg New York.
- Luisi, P.L. (2003) Autopoiesis: a review and a reappraisal. *Naturwissenschaften*, **90**: 49–59.
- Lunine, J.I. (1999) *Earth: Evolution of a habitable world*. Cambridge University Press, Cambridge.
- Madigan, M.T., Martinko, J.M. and Parker, J. (2002) *Brock Biology of Microorganisms*, 10th edn. Prentice-Hall, Englewood Cliffs, NJ.
- Malin, M.C. and Edgett, K.S. (2003) Evidence for persistent flow and aqueous sedimentation on early Mars. *Science*, **302**: 1931–1934.

- Martin, W. and Muller, M. (1998) The hydrogen hypothesis for the first eukaryote. *Nature*, **392**: 37–41.
- Moreira, D. and López-García, P. (1998) Symbiosis between methanogenic archaea and delta-Proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J. Mol. Evol.*, **47**: 517–530.
- Morse, S.S.E. (1993) *The Evolutionary Biology of Viruses*. Raven Press, New York.
- Murphy, F.A. (1995) Virus taxonomy. In: Fields B.N. and Knipe D.M. (eds.) *Fundamental Virology*. Raven Press, New York.
- Oren, A. (1994) The ecology of extremely halophilic archaea. *FEMS Microbiol. Rev.*, **13**: 415–440.
- Pedersen, K. (2000) Exploration of deep intraterrestrial microbial life: current perspectives. *FEMS Microbiol. Lett.*, **185**: 9–16.
- Price, P.B. and Sowers, T. 2004. Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. *Proc. Natl. Acad. Sci. USA*, **101**: 4631–4636.
- Reaney, D.C. (1974) On the origin of prokaryotes. *Theor. Biol.*, **48**: 243–251.
- Rivkina, E.M., Friedmann, E.I., McKay, C.P. and Gilichinsky, D.A. (2000) Metabolic activity of permafrost bacteria below the freezing point. *Appl. Environ. Microbiol.*, **66**: 3230–3233.
- Rothschild, L.J. and Mancinelli, R.L. (2001) Life in extreme environments. *Nature*, **409**: 1092–1101.
- Schleper, C., Puehler, G., Holz, I., Gambacorta, A., Janekovic, D., Santarius, U., Klenk, H.P. and Zillig, W. (1995) *Picrophilus* gen. nov., fam. nov.: a novel aerobic, heterotrophic, thermoacidophilic genus and family comprising archaea capable of growth around pH 0. *J. Bacteriol.*, **177**: 7050–7059.
- Searcy, D.G., Stein, D.B. and Green, G.R. (1978) Phylogenetic affinities between eukaryotic cells and a thermophilic mycoplasma. *Biosystems*, **10**: 19–28.
- Sharma, A., Scott, J.H., Cody, G.D., Fogel, M.L., Hazen, R.M., Hemley, R.J. and Huntress, W.T. (2002) Microbial activity at gigapascal pressures. *Science*, **295**: 1514–1516.
- Takami, H., Inoue, A., Fuji, F. and Horikoshi, K. (1997) Microbial flora in the deepest sea mud of the Mariana Trench. *FEMS Microbiol. Lett.*, **152**: 279–285.
- Varela, F.G., Maturana, H.R. and Uribe, R. (1974) Autopoiesis: the organization of living systems, its characterization and a model. *Curr. Mod. Biol.*, **5**: 187–196.
- White, D.C., Phelps, T.J. and Onstott, T.C. (1998) What's up down there? *Curr. Opin. Microbiol.*, **1**: 286–290.
- Yayanos, A.A. (2002) Are cells viable at gigapascal pressures? *Science*, **297**: 295.

9 Impact Events and the Evolution of the Earth

Philippe Claeys

9.1 Introduction

9.1.1 Terrestrial Craters

In July 1994, the twenty-one small fragments of comet Shoemaker Levy 9 struck Jupiter. Images of these collisions appeared almost live on the Internet and on television, (see <http://www2.jpl.nasa.gov/sl9/> for the event's image library). The solar system is indeed a violent place; the numerous craters of various sizes visible on the moon also attest to this as well. This satellite itself is most likely the result of the collision, about ~ 4.5 Ga ago between the young Earth and a Mars-sized planet (Canup 2004). Collision is a major process in planetary evolution; catastrophic asteroid or comet impacts have occurred and will continue to occur in the solar system.

Today approximately 170 impact craters are known on Earth (Fig. 9.1); and a couple of new ones are discovered every year. A regularly updated list of craters is found on the following website: <http://www.unb.ca/passc/ImpactDatabase/>. There, the craters are listed by name, size, location and many are illustrated by excellent images. The traces of many more collisions have been erased by geological processes such as tectonic, erosion, sedimentation, volcanism or the produced craters are hidden, buried under kilometers of younger sediments. Many more impact events have taken place in the oceans, which cover more than two thirds of the Earth's surface. The size of terrestrial craters varies between a few tens of meters to more than 200 km. Geology also influences the age and size distribution of the known impact craters. The majority is younger than 200 million years (Fig. 9.2). Small impact structures (< 5 km) erode rapidly and are thus under-represented, except for the most recent ones. Many craters lie on the old continental shields, such as Scandinavia, Canada or Australia, where they are not much affected by erosion or tectonics for extended periods. Many craters remain to be discovered in Africa or South America and of course at the bottom of the oceans, where so far no or little systematic searches have been carried out.

Some 25 craters ($\sim 15\%$ of the currently known population) have formed in the oceans, but always in less than 200 m water-depth, on the continental shelf (Dypvik and Jansa 2003). Plate tectonic activity has brought many to the

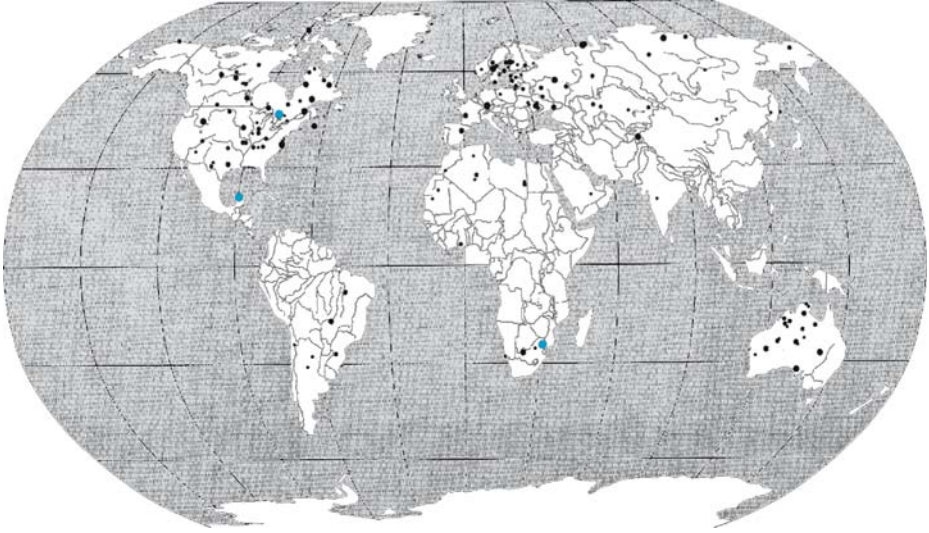


Fig. 9.1. Geographic distribution of known terrestrial craters; Chicxulub, Vredefort and Sudbury are in *blue*, the size of the *dot* is proportional to that of the crater (modified after French 1998)

surface and today only six of them remain in the marine realm. Their morphology and characteristics resemble those of craters formed on the emerged part of the continental crust (Dypvik et al. 2003). Several factors are responsible for the lack of identified oceanic impact structures. The ocean floors are young (< 200 million years old) and dynamic features, affected through plate tectonic activity by continuous production of fresh crust and subduction. Only a few large craters may have thus left a mark on the currently existing ocean floor. It is also extremely difficult to identify craters in the rare slabs of older oceanic crust, which are preserved, often strongly folded and/or metamorphosed, in the heart of mountain chains. Despite recent progress, the topography and cartography of the ocean floor remains too crude to allow the clear identification of an impact structure. The rapid sediment infill of a crater formed at the bottom of the ocean preserves its structure from erosion but at the same time quickly hides it from searches. It is also possible that the anatomy of a crater formed on the oceanic crust would differ from that formed of the continental crust.

Only one truly oceanic impact is presently known: it is the Eltanin event, which took place some 2.5 million years ago in the South Pacific Ocean, more precisely in the Bellingshausen Sea (Kyte et al. 1981; Margolis et al. 1991; Gersonde et al. 2002). A small ~ 1 km-diameter projectile fell in a 5 km-deep ocean and was completely melted and vaporized before reaching the seafloor; as a consequence no crater was formed. The traces of this event are limited to the debris ejected by the impact, and local evidence for major mass flows and transport

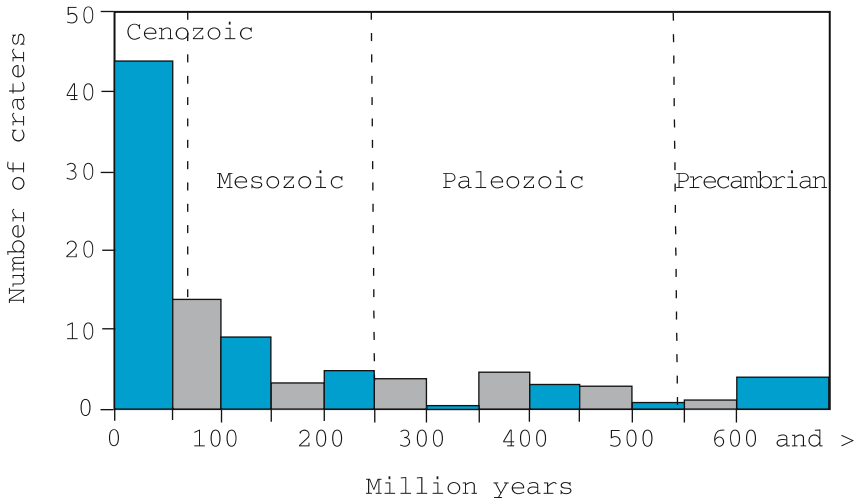


Fig. 9.2. Histogram showing the ages of terrestrial craters

of deep-water sediments. There is so far no example of the formation of an impact crater on the basaltic oceanic crust. Today, the oceans occupy two thirds of the Earth's surface and this proportion has been more important during parts of the planet's geological history. It must then be considered that impacts on the oceanic crust have taken place and that some high magnitude events have penetrated the ~ 7 km-thick oceanic crust. Their traces and consequences remain unknown. This leads to wild, hard-to-test but challenging speculations on the volcanic consequence of a crust-bursting impact event (see for example Jones et al. 2002; Ingle and Coffin 2004). However, according to Ivanov and Melosh (2003), an impact capable of perforating the crust and generating volcanism by decompression melting requires a projectile of ~ 50 km in diameter, which is possible but unlikely during the last 3.3Ga.

9.1.2 Historical Perspective of the Impact Process

The study of impact craters and their formation process have long remained a marginal topic for geologists. For many years, most Earth scientists interested in the formation of lunar or terrestrial craters preferred the volcanic hypothesis. This view was dominant until the second half of the twentieth century. It was difficult for many geologists to accept the cataclysmic effect of the collision of an extraterrestrial body with the Earth. Such a catastrophe did not fit well in the dogma of uniformitarianism established by the grandfather of geology, Sir Charles Lyell, as the scientific response to the "catastrophic view" of people like Baron Georges Cuvier. However, already in 1892, the famous American geologist Groove Karl Gilbert excluded any other origin than impact to explain the lunar cratering record. Alfred Wegener wrote in 1921 a short essay where he described

his experimental attempts of reproducing impact structures in the laboratory. In their excellent book on impact stratigraphy, Montanari and Koeberl (2000) drew an interesting parallel between the continental drift and impact theories. Both were developed early on by visionary researchers, probably ahead of their time, based on good common sense, solid evidence, and clear deductions, and both remained largely ignored if not scoffed at by the majority of geologists because they did not fit in the scientific paradigm of their time. In the case of impact, this view persisted until the 1970s despite unequivocal diagnostics of impact processes in the late 1940s (Dietz 1947). The vast majority of geologists had so many processes at hand to explain the morphology and shapes of impact structures that the far-fetched collision with an extraterrestrial object appeared superfluous and so much in contradiction to the so-often demonstrated principles of uniformitarianism (see discussions in Koeberl 2001).

In the 1960s and 1970s, the lunar exploration Apollo missions began to modify the negativism view most geologists had of impact processes. However, it was essentially the proposal made in 1980, that the fall of a ~ 10 km meteorite caused the extinction of the dinosaurs at the Cretaceous-Tertiary (KT) boundary (Alvarez et al. 1980; Smit and Hertogen 1980) that definitively established impact as a major geological process. The 1980s were punctuated by very critical and sometimes bitter debates about the impact-extinction scenario. Nevertheless, strong and unequivocal evidence accumulated in support of the impact theory; the story of this scientific adventure is pleasantly summarized by Alvarez (1997). In 1990, a large crater, named Chicxulub, was identified buried under the Yucatan peninsula in Mexico and dated precisely at the KT boundary (Hildebrand et al. 1990; Swisher et al. 1992). Today, the role played by the Chicxulub impact in the KT boundary mass extinction is widely accepted. Most geoscientists now recognize impact events as significant geological processes, which finally replaces the Earth in the global evolutionary perspective of the whole solar system.

9.2 Characteristics of Impact Craters

9.2.1 Magnitude and Frequency

Impact craters are topographical features commonly identified on planets and moons that have solid surfaces. A fresh crater is characterized by an almost perfectly circular morphology. The surface of the Moon is riddled with nicely rounded impact craters. Only a few structures display an elongated morphology. They result from very oblique impacts when the projectile hits the surface with an angle $< 10^\circ$. On Earth, tectonic, sedimentation or erosion activities can modify or even erase the original circular morphology, as is the case for example at the Sudbury crater in Canada (Deutsch et al. 1995).

The formation of an impact crater by collision with an object at hypervelocity is a complex process that shows similarities with nuclear explosions (Roddy

et al. 1977; Melosh 1989). The energy released by the impact results from the transformation of the kinetic energy of the projectile and is given by $1/2mv^2$, where m and v are respectively its mass and velocity. If the velocity reduction caused by the atmosphere is neglected, which is justified for decametric or larger objects, the Earth surface is struck with a velocity between 11 and 72 km/s. These velocities are determined by celestial mechanics. Eleven km/s is the escape velocity of the Earth-Moon system, a body with zero velocity is simply accelerated to this speed by a gravitational pull. The 72 km/s is the maximum estimated for a comet, it corresponds to the escape velocity of the solar system, at a distance of 1 AU, added to the average orbital velocity of the Earth (or $42 + 30$ km/s). On average, an asteroid strikes the Earth at a speed of ~ 20 km/s, and a comet at a much higher velocity around 60 km/s (Chyba et al. 1994). These collisions thus release enormous amounts of energy, superior, even for small bodies (~ 50 m) to that of a nuclear bomb (Table 9.1). Over a time frame of a few seconds, and over a limited surface, the impact of a 1.5 km-sized object, a size smaller than many known asteroids, releases approximately the same amount of energy ($\sim 1.3 \times 10^{21}$ J) as the whole internal Earth system (heat flow, earthquakes, tectonic, volcanism, etc.) during an entire year.

The estimations of the flux of objects impacting the Earth vary according to different authors and the approximation method selected (Shoemaker et al. 1990; Bottke et al. 1994; Grieve and Shoemaker 1994; Neukum and Ivanov 1994). Several approaches are used to estimate this frequency: scaling up the flux of very small objects (micrometeorites) penetrating in the atmosphere; extrapolation based on the number and size of craters preserved on the surface of other planets, in particular the moon; tracking the orbits and number of known asteroids and comets and counting the number of craters on Earth and their stratigraphy position. Table 9.1 provides a conservative estimation of

Table 9.1. Estimation of the collision frequency

Size of projectile	Frequency (1/...)	Example and released energy (eq. TNT)
< 50 m	Frequently detected	Burns in atmosphere
50 m	200 to 400 years	Tunguska ~ 12 MT
100 m	1000 to 5000 years	Wolf Creek, Pretoria, Barringer ~ 15 MT
500 m	0.1 to 0.5 Myr	Zhamanshin, Bosumtwi, Mien $\sim 11 \times 10^3$ MT
1 km	1 Myr	El Gygytgyn, Ries, Rochechouart $\sim 9 \times 10^4$ MT
5 km	10 to 50 Myr	Popigai, Manicouagan $\sim 1 \times 10^7$ MT
10 km	100 to 500 Myr	Chicxulub, Sudbury, Vredefort $\sim 1 \times 10^8$ MT
> 10 km	Probable?	Precambrian ejecta in Australia and Africa?
> 100 km	Probable?	Late Heavy Bombardment
> 1000 km	At least 1	Origin of the Moon

the collision frequency; it is based on French (1998), where a more complete discussion can be found. The following website: http://www.lpl.arizona.edu/SIC/impact_cratering/World_Craters_Web/Chronologypage.html gives the stratigraphic distribution of terrestrial craters. On Earth, there is a 1/20 ratio between the size of a projectile and that of the resulting crater. For small objects (< 100m) the composition and its mechanical strength are also important. A stony meteorite less than 50m in diameter does not survive its passage through the atmosphere; it disaggregates and/or explodes before touching the ground. The vast majority of the small < 1.5km craters on Earth were thus produced by iron meteorites.

To this day, asteroid or comet impacts have never been directly observed on Earth, despite possible speculations regarding mythological events. In 1908, a small object, maybe a stony meteorite ~50m in size, exploded several kilometers in the atmosphere above Siberia. The Tunguska event, as it is called, devastated more than 2000km² of forests. Its entry in the atmosphere over a large city would have wiped it off the map. Current knowledge of the impact process is thus based on the study of preserved craters, planetary observation by remote sensing, laboratory experiments and mathematical modeling. The high pressures and temperatures generated during an impact event significantly affect the crustal lithologies at the point of collision (“the target rock”); the characteristic modifications induced render the identification of ancient impact structures possible.

9.2.2 Crater Morphologies

An impact crater is in theory the cavity excavated by the impact; it must be distinguished from an impact structure, which is the result of geological modifications of the original morphology with time. Most commonly, both terms are used indistinctively. Several crater morphologies are known on Earth and the other rocky planets (Fig. 9.3). A simple crater has a well-rounded bowl-shape, and is characteristic of craters between < 2km in diameter to maximum 4km, depending on the impacted lithologies. Its depth is equivalent to approximately 1/3 of its diameter. Above 4km in size, a complex crater forms. It is marked by a central uplifted part and its concavity decreases. The depth of such crater is equivalent to about 1/6 or less of its diameter. During the impact, the rocks in the crater center were compressed, the central peak forms through the uplift of the deep lithologies as the pressure is released. The center of the depression contains melted and highly shocked lithologies formed by the high temperatures and pressures generated during the impact. These characteristic lithologies (called impactites) rest on a highly fractured basement. In general, a pool of melt rock is overlain by suevite, which is a clastic breccia containing melt-fragments. An elevated rim limits the depression and the inner flanks contain lithic breccia, fractured rocks, debris slope material, and eventually patches of melt rock and/or suevite.

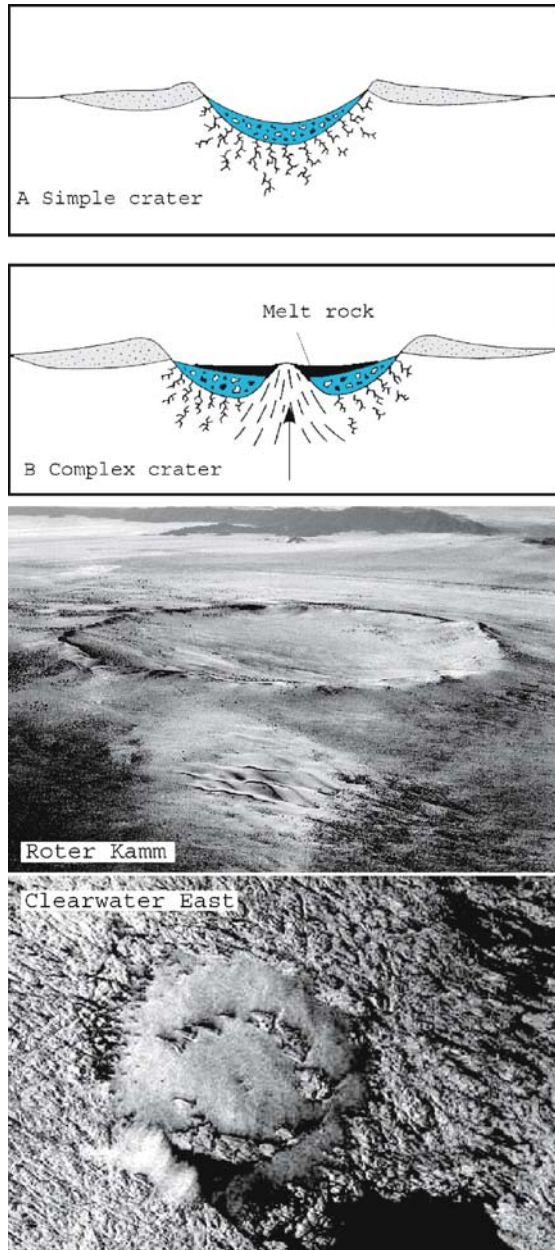


Fig. 9.3. Impact crater morphologies, simple bowl-shaped crater (**A**) and complex crater with central peak (**B**) (modified after French 1998), illustrated by photos from the Roter Kamm (Namibia, 2.5 km) and Clearwater East (Canada, 26 km) (Roter Kamm, photo courtesy of Christian Koeberl, Clearwater: photo Earth Impact database, www.unb.ca/passc/ImpactDatabase)

Craters larger than a few hundred kilometers present a complex and still poorly understood morphology composed of several elevated concentric rings. They are called multiring craters. Structures of this type reaching more than 2000km in diameter are known on the Moon, Mercury, Mars, Callisto, and Ganymede (Spudis 1993). It is even possible that the three largest terrestrial craters, Chicxulub (Mexico, ~200km), Sudbury (Canada ~250km) and Vredefort (South Africa ~300km) present this multiring morphology. Unfortunately, their internal structure cannot be clearly identified; the first one is buried under 1km of younger sediments and the other two that are Proterozoic in age, are strongly tectonized and eroded.

9.2.3 Formation Mechanism
(Based Essentially on Melosh 1989 and French 1998)

A crater forms by propagation of intense shock waves generated at the contact point between the projectile and the target lithologies (Fig. 9.4). These waves propagate at velocities >10km/s, and are transient stress waves, reaching pressures above 100GPa, which differ from waves produced by other geological

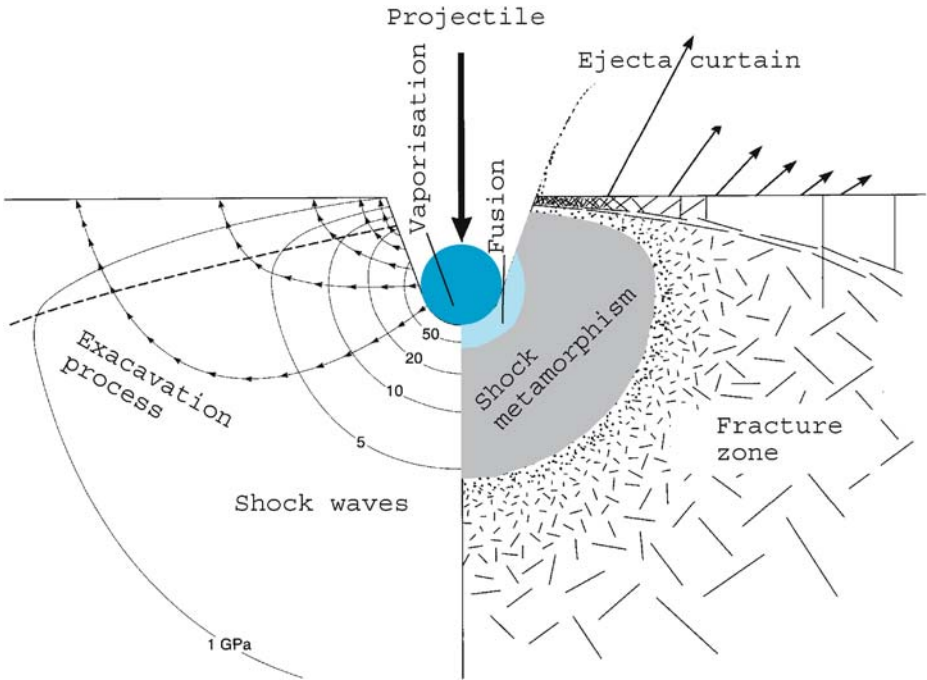


Fig. 9.4. Formation of a crater and propagation of the shock waves; the zones where vaporization, fusion and shock metamorphism are indicated (modified after French 1998)

processes such as earthquakes. Terrestrial rocks deform elastically or plastically around ~ 1 GPa. The pressures generated are superior to this stress level and shock waves generate characteristic modifications and deformations in the rock through which they propagate. The passage of the shock wave at high velocity ($>$ the speed of sound in the crust) vaporizes, melts, deforms, compresses and shoves a large volume of rocks leading to the excavation of a cavity (Fig. 9.5). At this scale, such a process cannot be reproduced in the laboratory; its understanding is derived from theoretical calculations, laboratory experiments on the behavior of shock waves in rocks, the study of mini-craters obtained by shooting targets of specific compositions, and field observations at preserved impact structures. In present knowledge, this extremely rapid and continuous process can be divided into three phases (see Melosh 1989, or French 1998 for a more elaborate discussion): (1) contact and compression, (2) excavation, and (3) post-impact modifications (Fig. 9.5).

During the first phase, the projectile penetrates a solid target down to a maximum depth of twice its radius. An enormous amount of kinetic energy is transferred to the target rocks by the propagation of shock waves originating at the contact point (O'Keefe and Ahrens 1993). Several aspects of the conversion of the kinetic energy into shock waves remain poorly understood. A front of shock waves expands outward into the surrounding target rocks while another set of waves is reflected into the projectile. In the target, the shock waves quickly lose energy because of the continuously increasing volume they cover and through heating, deformation and excavation of the rocks. Their peak pressures decrease exponentially with distance from the impact point and can be represented by a law of the R^{-n} type, where n depends on the projectile size and impact speed ($-2 > n > -4.5$) and R represent the distance to the contact point (Ahrens and O'Keefe 1977). At the distance where the velocity becomes equivalent to that of sound in rocks (5–8 km/s), the shock wave transforms into a seismic wave. In the case of a large impact, the latter can be felt all over the planet. The contact point is thus surrounded by a series of concentric hemispherical shock zones, of diminishing pressure intensities (Fig. 9.4). The shock wave compresses the lithologies up to pressure above the Hugoniot elastic limit, generally defined as the maximum constraint that a material can withstand without deforming permanently and situated for most minerals and rocks between 5 and 10 GPa. After the passage of the shock wave, the generated pressure is released by a rarefaction wave that follows and catches up on the shock wave. During the compression, an enormous amount of energy is injected in the target rock and in the projectile. During the decompression, the excess heat released leads to their vaporization and/or melting. This results in the almost complete vaporization of the projectile and a large part of the target rock.

The vaporized projectile plays no role in the excavation phase. At the maximum penetration point, complex interactions take place between shock and rarefaction waves, which push the target rocks and set it in movement. Fluxes of material are displaced in various directions (Figs. 9.4 and 9.5). In the upper part

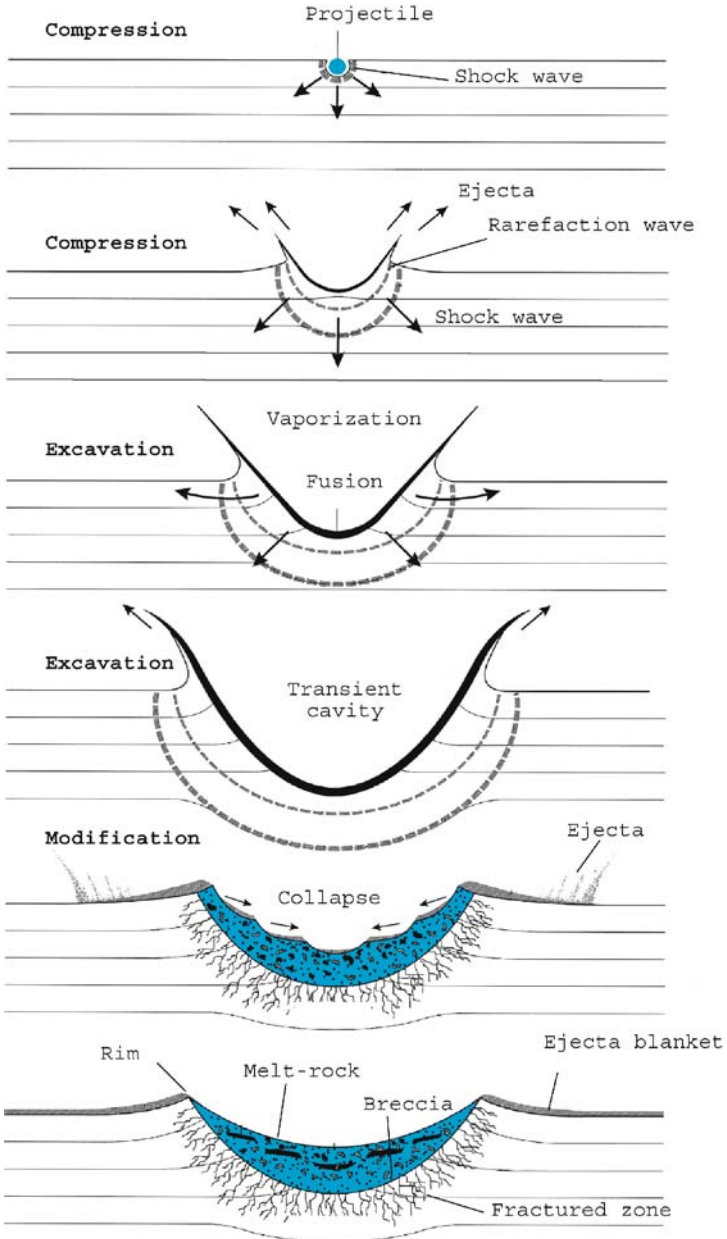


Fig. 9.5. Formation process of a simple crater; in the central peak of the crater the end of the excavation phase would be marked by the uplift of deep lithologies in the center (modified after French 1998)

of the depression created by the projectile, the rocks are displaced in an upward direction with such velocities (several km/s) that a significant part is expelled from the crater. At the vertical of the contact point, a large cloud of vapor, melt and dust grows above the forming crater and rises into the atmosphere. Target rock fragments are also ejected on more oblique trajectories in the form of a curtain of ejecta. The ejecta curtain material that is composed of melted, shocked, and solid fragments will land at a distance of several crater radii from the rim. Even at the periphery of the cavity, the velocities are high enough to expel large blocks, which deposition beyond the edge of the crater forms a thick ejecta blanket. The cavity expands thus to a size much larger ($\sim 20x$) than that of the projectile. In the deepest part of the newly formed cavity, the fractured lithologies are pushed downward and/or outward towards the sides, but in most cases, they remain within the depression. The displacement of material leads to the formation of a bowl-shaped depression, called the transient cavity. This excavation phase continues as long as the expanding shock wave is energetic enough to displace or eject the target rock.

The last phase starts as soon as the transient cavity has reached its maximum size. Gravity and rock mechanics take over leading to the collapse of the crater rim and other unstable zones, filling the depression with breccia and large blocks. The Barringer crater in Arizona (1 km) formed in a few seconds, whereas a much larger crater, such as Vredefort in South Africa, may have taken up to 10 minutes. At this point, the cratering phase has theoretically ended; however, various modifications can still take place over long periods of time due to tectonic activities (fracturing, or earthquakes), sedimentation or erosion.

9.2.4 Identification Criteria

On inactive planets without an atmosphere, their circular morphology facilitates the identification of impact craters, but they remain inaccessible except by remote sensing. On Earth, where field observation is directly possible, tectonic activity coupled with metamorphism, sedimentation and erosion processes can significantly complicate the identification and study of impact craters. A set of criteria enables the identification of impact craters. These decisive factors are: (1) clear geophysical anomalies, (2) evidence for shock metamorphism in minerals, and (3) the presence of preserved meteorite fragments, or most frequently a geochemical contamination by a meteoritic component.

Gravity measurements, magnetic properties and seismic analyses can be used to outline the morphology and shape of impact craters (see Personen and Henkel 1992; Pilkington and Grieve 1992; Henkel and Reimold 1998, for more details). A local negative anomaly of the gravity usually outlines the morphology of a simple crater. It is generated by the relatively lower density of breccia and highly fractured rocks present within the structure compared to the surrounding lithologies. A complex crater is characterized by a central positive gravity anomaly

due to the presence of denser lithologies in the central peak surrounded by lower gravity in the trough zone, where the rocks are highly fractured and brecciated. This central peak zone may also be marked by a positive magnetic anomaly caused by the presence of uplifted iron-rich rocks. Seismic profiles document the structural deformation and often outline the general bowl-shaped silhouette of the crater cavity, allowing in some cases a precise estimate of its size. Coherent reflectors are absent from the central zone of the depression, while on the flank highly fractured lithologies and collapse along a slope of large blocks can induce a loss of signal quality and coherence. Seismic studies make it also possible to distinguish a buried crater from a volcanic structure based on the absence of roots connecting it to the lower crust or mantle. Even large craters remain relatively shallow crustal structures, the deep reflectors extend undisturbed below the cavity. Geophysical exploration has led to the discovery of numerous craters, especially when buried under water of younger sediments (Tsikalas et al. 1998 for example). Such detailed geophysical explorations have made Finland the most cratered region in the world (Pesonen 1996). It is in large part thanks to geophysical studies that the famous Chicxulub crater buried under ~1km of Cenozoic sediments was discovered at the tip of the Yucatan peninsula (Penfield et al. 1981; Hildebrand et al. 1991; Pilkington et al. 1996). However, geophysical measurements alone are unable to categorically identify a structure as an impact crater.

Evidence of shock metamorphism in minerals is a clear and diagnostic factor to identify an impact crater. At the macroscopic scale shock metamorphism is demonstrated by the presence of shatter cones (Fig. 9.6) and at the microscopic scale by permanent modification of the crystallographic texture of minerals, of which some of the best examples are found in quartz. The formation mech-



Fig. 9.6. Shatter cone from Sudbury, Canada

anism of shatter cones remains poorly understood. The passage of the shock wave confers to the rocks a distinctive striated pattern of fractures in the form of a cone from millimeters to meters in size. They are best visible in fine grain lithologies such as limestone (Rondot 2003). As a dynamic shock wave passes through minerals, it modifies irreversibly their crystallographic texture, leading to the formation of microscopic defects. In quartz, the most characteristic and diagnostic evidence of shock metamorphism is the presence of the planar deformation features (PDFs) (Fig. 9.7). The presence of its high-pressure (> 10 GPa) polymorph stishovite is also a diagnostic criteria of shock metamorphism. The PDFs form as a network of fine lines oriented parallel to two (or more) specific crystallographic planes of the mineral. Goltrant et al. (1991) have demonstrated using a transmission electron microscope that these lines are in fact very narrow lamellae ($0.1\ \mu\text{m}$) of amorphous SiO_2 . The passage of a shock wave with a pressure above 5 GPa causes the localized amorphization of the crystalline structure (recent reviews are found in Stoeffler et al. 1994; Doukhan and Leroux 1996; French 1998). This passage is too rapid to break the chemical bonds or to allow the atoms to reorganize under a new crystallographic structure. If the pressure of the wave is much higher, the whole quartz crystal is completely annealed. In contradiction to some discussions, which resurface now and then in the literature (e.g., Lyons et al. 1993), the pressures generated by tectonic or volcanic processes on Earth are several orders of magnitude lower and unable to produce such deformations of the crystallographic structure. The required dynamic shock

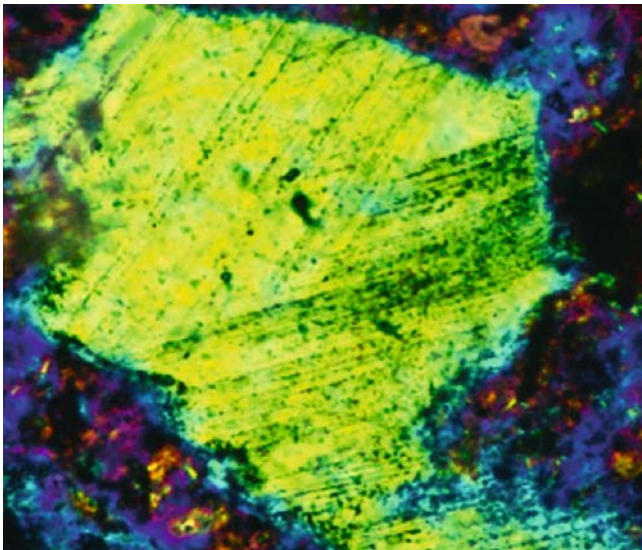


Fig. 9.7. Planar deformation feature in quartz clearly showing two directions (Manson Crater, Iowa, USA, photo courtesy of Christian Koeberl, University Vienna). The grain is approximately 150 micrometers

pressure of at least 5 GPa is never reached even in the most intensely tectonized or metamorphosed zones. The only other known process capable of inducing shock metamorphism in minerals is a nuclear explosion.

In rare cases, meteorite fragments are preserved at close proximity to the crater (e.g., Barringer in Arizona). In the absence of such clear evidence, it becomes necessary to search for a possible meteoritic contamination in the impactite lithologies. Siderophile elements such as Cr, Co, Ni, but most of all the platinum group elements (PGEs: Ir, Pt, Pd, Rh, Ru, Os) are more abundant in non-differentiated meteorites than in the Earth's crust (Fig. 9.8). In the case of the PGEs, this enrichment reaches several orders of magnitude. The presence in the melt rock or in the ejecta material of enrichment in PGE is thus a clear signal of an impact. Chemical analyses using neutron activation or inductively coupled plasma mass spectrometry can reveal even a minute meteoritic contamination (less than 1 wt%) by these elements. By comparing the elemental ratios of the different PGEs, it is even possible in some cases to characterize the type

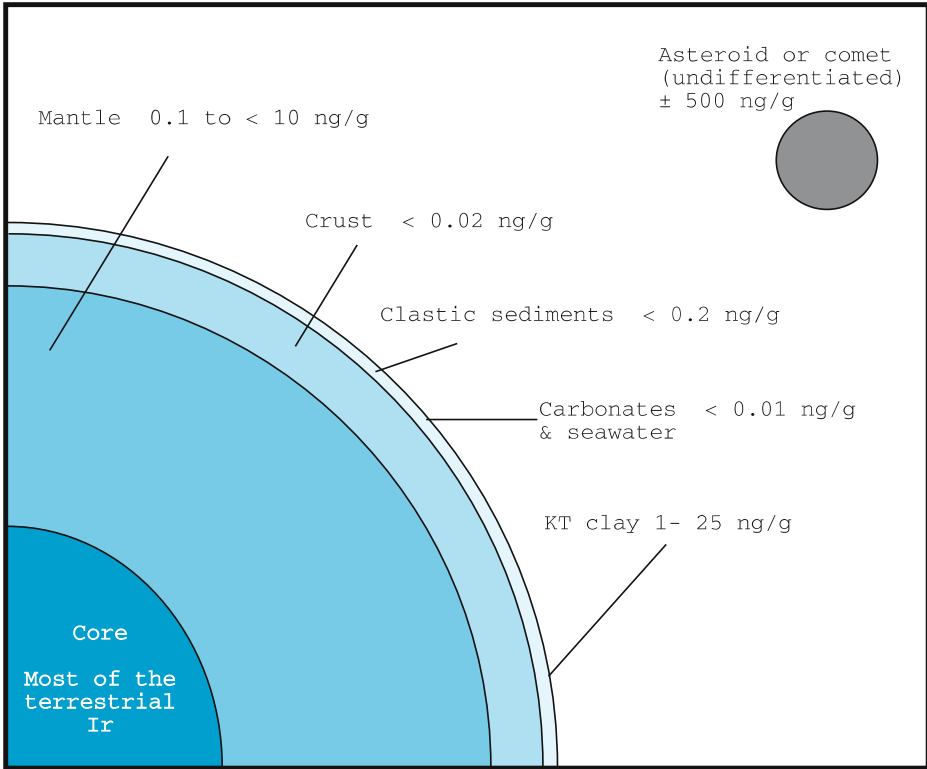


Fig. 9.8. Distribution of Ir on Earth (representative of all the PGEs) in ng/g or ppb (10^{-9} g/g or part per billion). Most PGEs are most likely concentrated in the Earth's core

of impacting meteorite, for example to distinguish between an iron or different types of chondrites (McDonald et al. 2002; Tagle and Claeys 2004). The isotopic ratios of Os and Cr differ significantly in meteorites and in the Earth crust. Measurements of the $^{187}\text{Os}/^{188}\text{Os}$ and/or $^{53}\text{Cr}/^{52}\text{Cr}$ ratios of the melt rock also provide a clear indication of a meteoritic contamination (see extensive review by Koeberl et al. 2002).

The presence in the ejecta material of cosmic spinels is a mineralogical argument indicative of an impact event (Robin et al. 1992). Spinel is an extensive family of secondary minerals represented in terrestrial rocks mainly by magnetite and chromite. The impact-related spinels are enriched in Ni (up to 6 wt%) and most of their iron (80 to 90 per cent) is in an oxidized (3^+) state. Their counterpart found in sediments or volcanic rocks, rarely contain more than 1 wt% Ni and their Fe is systematically in a reduced (2^+) state. These oxidized Ni-rich spinels are called magnesioferrite and/or cosmic spinels. Their formation results from the fusion and oxidation of the meteorite during its entry and course through the atmosphere (Robin et al. 1992). According to other authors, they may also form directly by condensation out of the impact induced vapor cloud (Kyte and Bostwick 1995).

Another criterion to recognize an impact event is the presence at the periphery of the crater of glass fragments, or at further distances the presence of tektites or microtektites. Impact glasses are found surrounding small craters such as Aouelloul in Mauritania or Darwin in Australia (Koeberl et al. 1998). They unambiguously differentiate from volcanic glasses by their low water content ($< 0.1\text{ wt}\%$) and their chemical composition reflecting that of impacted lithologies. Tektites (and microtektites) are centimeter (or millimeter) sized rounded or elongated glass particles of homogeneous composition and rich in SiO_2 (Fig. 9.9). During an impact event, they form by fusion of the uppermost part of the target rock, are then ejected at high velocity from the crater, then quenched in flight,



Fig. 9.9. Tektites from the most recent strewnfield in South East Asia

which explains their morphologies, and finally deposited over a vast geographic area called a “strewnfield”. The exact formation and ejection processes early on in the cratering event remain unfortunately poorly understood. The chemical composition of tektites reflects that of the upper part of the target rock at the impact site. They can be found at a distance of several thousand kilometers from their source crater. Four strewnfields are known in the Cenozoic, in chronological order: (1) the North American (~ 35 Ma) originating from the Chesapeake Bay crater offshore Maryland and Virginia, (2) the Central European (~ 15 Ma) ejected by the Ries impact in Bavaria all the way to the Czech Republic, (3) the Ivory Coast (~ 1 Ma), which originates from the Bosumtwi crater in nearby Ghana, and (4) most recently the Australasian (~ 0.7 Ma) extending from all over the Indian ocean from Taiwan to Madagascar and from Southern India to Australia (see Koeberl 1986; Montanari and Koeberl 2000 for a more detailed discussion).

9.3 Case Study: The Cretaceous-Tertiary Boundary and the Chicxulub Crater

The Cretaceous-Tertiary (KT) boundary event took place 65Ma ago; it must be considered as a case study to understand the role of impacts in the evolution of the Earth, the dynamics of crater formation, as well as the production and distribution of ejecta material. It represents the only well documented stratigraphic horizon where an impact event is linked to a major biological crisis. The KT boundary mass extinction led to the demise of the dinosaurs and approximately 50% of the Earth’s fauna and flora including marine reptiles, calcareous microorganisms and invertebrates such as the ammonites. It is difficult to identify a clear pattern of selection among survivors and victims. Apparently, small, perhaps omnivorous, versatile animals were favored, but almost all those weighing more than 25kg disappeared (Buffetaut 1996). Several articles have discussed the KT boundary problem. They start with the discovery of the now famous Ir positive anomaly in the boundary clay in Italy and Spain (Alvarez et al. 1980; Smit and Hertogen 1980). They continue with the recognition of the buried impact Chicxulub crater under the Yucatan platform (Camargo and Penfield 1980; Hildebrand et al. 1990), and finish with the presentation of the latest drilling results at Chicxulub (Urrutia et al. 2004). This chapter does not follow the chronological order of the discoveries; rather it begins with the crater description, then examines the distribution of ejecta, and at the end the consequences of the impact for the climate and biosphere are discussed in more detail.

9.3.1 The Chicxulub Crater

Chicxulub can be considered as a *model* crater, as for a structure of this size (~ 200 km) it is young and little affected by alteration or tectonic processes. Its

rapid burial by Cenozoic sediments contributes to its preservation but also limits its study to geophysical approaches and sample recovery by deep drilling. High-precision measurements of the regional gravity field clearly outline the central peak ring, surrounded by a region of lower gravity caused by low-density lithologies present in the trough zone (Pilkington et al. 1994). A magnetic anomaly induced by the presence of iron-rich lithologies closely coincides with the gravity anomaly marking the central peak ring. The deep crustal materials compressed by the impact and uplifted towards the center of the structure probably contain more iron than the sediments, which form the dominant lithology of the Yucatan peninsula. Seismic profiles, obtained in the offshore part of the structure, delineate its concave morphology and indicate a transient cavity between 80 and 110km in size (Morgan et al. 1997). Final size estimation varies depending on the interpretation of the structure as a central peak ring or a multiring crater (see Morgan et al. 2002 for discussion). A diameter of approximately 200km is considered a reasonable size estimate for the Chicxulub crater. However, geophysical imaging alone is not sufficient to definitively demonstrate the crater origin of a deeply buried structure: sample examination is required.

The cores, obtained more than 35 years ago for oil exploration purposes by the Mexican oil company Pemex, coupled with the recent results of the Chicxulub Deep Drilling Program of the International Continental Scientific Drilling Program (ICDP), offer a view of the impactites present within the structure. The University of Mexico City (UNAM) also drilled the crater margin in 1994. Inside the Chicxulub structure, underneath more or less 1km of Cainozoic sediments, the cores reveal the presence of suevite, melt rock and breccias (Fig. 9.10), resembling the impactite sequences found at other craters such as the Ries in Germany, or Sudbury and Manicouagan in Canada. In the Pemex Yucatan 6 well for example, the suevite—a chaotic breccia rich in formerly melted particles—overlies a characteristic melt rock. The latter is formed by partially or totally melted target rock fragments floating in a rapidly cooled glassy or cryptocrystalline matrix. Most of the Yucatan platform original lithology can still be identified as shattered fragments in the Chicxulub suevite breccia. In the suevite, the basement fragments often contain shocked quartz, seldom shocked feldspar and zircon crystals. The evidence for shock metamorphism in quartz such as the orientation of the PDF and presence of diaplectic quartz, where the grain was completely melted by very high-pressure shock waves, indicate that these rocks underwent dynamic pressure in the range between 10 and 65 GPa. The fine matrix present between the basement clasts in the melt rock has been dated by the $^{40}\text{Ar}/^{39}\text{Ar}$ and yielded an age of $65.46 \pm 0.6\text{Ma}$ (Swisher et al. 1992, recalculated by Renne et al. 1998). This age is perfectly compatible with that commonly accepted for the KT boundary. Moreover, paleomagnetic studies indicate that the Chicxulub melt rock solidified under a period of reverse magnetic polarity (Sharpton et al. 1992). In sedimentary sequences worldwide, the KT boundary clay occurs in the same reverse magnetic sequence labeled 29R. The formation

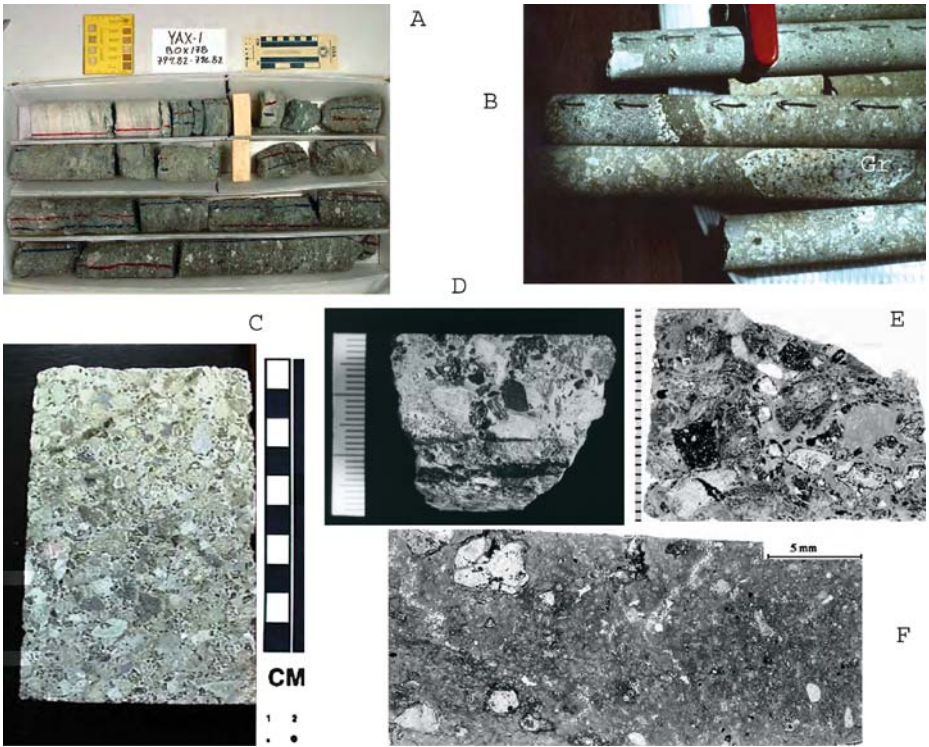


Fig. 9.10. Different types of impactites from the Chicxulub crater (A) top of the sequence of suevite passing to sediment of basal Palaeocene (Yaxcopoil ICDP well, 795 m). (B) Polymict breccia containing clasts of the Yucatan basement (*Gr* = granite) carbonate and evaporite are also present (UNAM-6 just outside crater rim). (C) suevite rich in carbonate and glass fragments now altered to greenish clay minerals (Yucatan 6 around 1190 m). (D) suevite containing essentially deep basement clasts and glass (*black*) (Yucatan 6 around 1225 m). (E) thin section at the same level as D showing the mixing of clasts originating from different depth-levels in the target rock. (F) melt rock (Yucatan 6 around 1280 m)

of the Chicxulub crater and the deposition of the KT boundary clay layer do coincide.

In the Late Cretaceous, the Yucatan platform consisted of a 3km-thick sequence of carbonates, interlayered with evaporites (CaSO_4) deposited on a shallow water platform, overlying a silicate-rich Pan African age (~550Ma) basement, whose overall composition was most likely granitic. The suevite layers recovered from the Pemex and ICDP drilling programs (-1100m in Yucatan 6 and -840m in Yaxcopoil 1) are often rich in CaCO_3 . The vaporization and fusion of the lithologies present in the upper part of the Cretaceous Yucatan platform have thus played a key role in the impact event

by releasing a large volume of gases in the atmosphere. The precise estimation of the quantities of CO₂ and sulfur component injected in the atmosphere is important to understand the consequence of the impact for the climate and the biosphere (see discussion in Jones et al. 2000; Claeys et al. 2003).

9.3.2 Distribution of Ejecta

A continuous ejecta blanket covers most of the Yucatan platform extending to more than 360km from Chicxulub. It was deposited by the huge mass of material, consisting of large blocks, vapor and melt fragments expelled from the crater. This tremendous erosive and transport capability of this high-energy debris flow was capable of eroding and displacing the uppermost layers of the Yucatan platform (Fig. 9.11). Just outside the crater rim, the results of the UNAM shallow drilling campaign shows that the proximal ejecta blanket is formed by several hundred meters of polymict breccia containing carbonates, evaporites and melted or solid fragments originating from the deep basement under the crater. The proportion of crater material clearly decreases with increasing distance from the rim. Near the Mexico-Belize-Guatemala



Fig. 9.11. Ejecta blanket in Belize, at some 360km from Chicxulub. The stratification is clearly visible in this large block that was eroded from the upper part of the Yucatan platform, transported, and incorporated as one of the many fragments of the breccia



Fig. 9.12. Within the breccia of Fig. 9.11, small (cm-sized) fragments of basement clasts are present; they are now composed of clay minerals but most likely originated from the deeper lithologies underneath the Yucatan (~5 km). Most likely, they fell from the periphery of the cloud of dust in gas that was expanding over the crater and were incorporated to the moving ejecta blanket. It is also possible that such fragments were already present in the mass of rocks that was expelled from the crater. In the ejecta blanket the proportion of crater material decreases with increasing distance from the crater

borders, the ejecta blanket has thinned down to a 20-meter-thick monomineral breccia essentially formed of locally eroded material. It is composed of chaotic blocks of the Upper Cretaceous Barton Creek dolomite, ranging in size from a few millimeters to several meters surrounded by a matrix of finely crushed dolomite and clay minerals with rare highly altered basement clasts (Fig. 9.12).

All around the Gulf of Mexico, from the region south of Monterey, to Texas and Alabama, as well as in Haiti, the KT boundary is marked by a coarse clastic sequence several meters thick (Smit 1999). In the late Cretaceous, this whole zone was under several hundred meters of water. This complex sedimentation sequence (Fig. 9.13) is sandwiched between fine marls of the Upper Cretaceous and Lower Tertiary age: the typical low energy, fine-grained sediment expected in such a deep-water marine environment. The widespread distribution all over the Gulf of Mexico of a coarse clastic sedimentation in deep water is unusual. It can only be explained in the context of the formation of the nearby Chicxulub crater. The basal unit of the sequence that erodes the Uppermost Cretaceous deep-water sediment contains elongated or rounded millimeter-sized clay spherules, shocked quartz and allochthonous carbonate fragments. The carbonate grains originate from shallow water facies, similar to those that must have existed on the Yucatan platform in the Upper Cretaceous. Typical impact glass with very

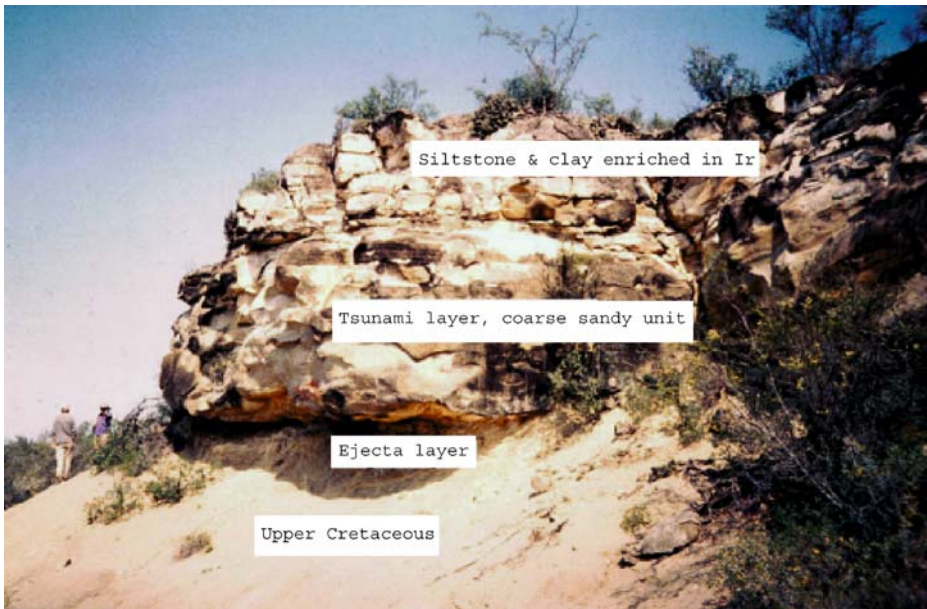


Fig. 9.13. KT boundary at Rancho Nuevo (NE Mexico) is marked by a thick clastic sequence of ejecta, coarse laminated sandstones and fining upward clay and silt units. The succession of these units illustrates from base to top, the arrival of the ejecta curtain, followed by the sandy material eroded and transported by the tsunami waves, and at the end the finer sediments with the clay units of enriched in PGE. Such sequence is found all over the Gulf of Mexico, from Alabama to NE Mexico all the way to Haiti

low water content (Koeberl 1992) is found preserved in the core of some of the aerodynamically shaped spherules (Fig. 9.14). The chemical and isotopic compositions of the glass are identical to those of the Chicxulub melt rock. The radiometric age of the impact glass obtained also by the $^{40}\text{Ar}/^{39}\text{Ar}$ method is indistinguishable from that of the Chicxulub impact melt, clearly indicating their common origin (Swisher et al. 1992). These particles were deposited by the ejecta curtain expelled from the crater along ballistic trajectories with take-off angles of 30 to 45°. A one to two meters-thick well-laminated sandstone layer covers this ejecta unit. This layer is interpreted as caused by the repetitive passage(s) of huge tsunami wave(s). Upon breaking on the continent, these megawaves eroded huge volumes of coastal sand, and as the water pulled back, this coarse material was entrained and deposited far offshore in deep water. The whole sand unit was most likely deposited in a short time period, perhaps several days to a few weeks. The tsunamis were generated by the huge displacement of material induced by the formation of the crater and the large earthquakes, which magnitude may have reached 13 on the Richter scale according to the calculations of O'Keefe and Ahrens (1991). The sandstone layer fines upwards (probably as the megawave decreased in energy) and grades to alternating layers



Fig. 9.14. Spherules and impact melt glass found in the basal unit of the KT boundary sequence in Northeastern Mexico. The chemical composition, isotope signatures and ^{40}Ar - ^{39}Ar ages are similar to those measured in melt rock within the Chicxulub crater. The black glass (*bottom left*) with an andesitic composition ($\sim 65\text{wt}\%$ SiO_2) resulted from melting the silicate-rich lithologies deeply buried underneath Yucatan; the yellow/orange glass (*bottom right*) contains up to 28wt% CaO was produced by mixing the molten basement rocks with the overlying carbonates

of silt and clay (Fig. 9.13). The clay is systematically enriched in PGE and Ni-rich spinels. The base of the Palaeocene is found directly above this third unit.

A thin (a few centimeters) clay layer marks the KT boundary over the rest of the world (Fig. 9.15). The Ir anomaly that was first detected in this layer in Italy and Spain is now identified at 112 sites worldwide (Fig. 9.16; Claeys et al. 2002). It is in most sites accompanied by dendritic crystals of Ni-rich magnesioferrite spinels (Fig. 9.17), and shocked minerals, in particular quartz, which was a decisive element to support the impact hypothesis (Bohor et al. 1987). These fine sediments located at distal sites from the Chicxulub crater are the result of the condensation and sedimentation of the materials present in the huge vapor cloud that rose above the crater and quickly engulfed the whole planet after the impact. Models indicate that this cloud of vapor and fine dust took about a year to dissipate. Such a high concentration of dust in the upper atmosphere is believed to have signifi-

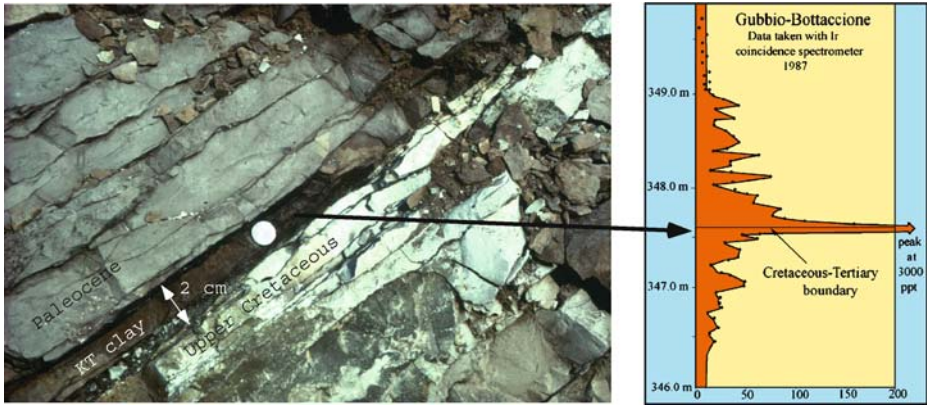


Fig. 9.15. KT boundary at Contessa (Italy), the Ir profile shown on the *right side* of the figure was originally measured by Alvarez et al. (1980) in the nearby Gubbio section

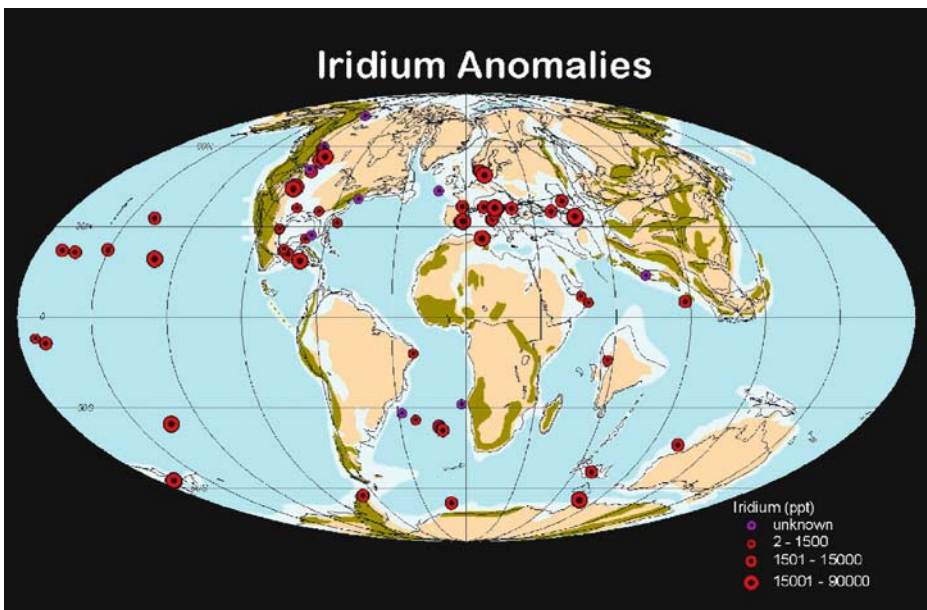


Fig. 9.16. Paleogeographic map of the Late Cretaceous displaying the KT sites where a positive Ir anomaly has been identified, the size of the dot is proportional to the Ir concentration

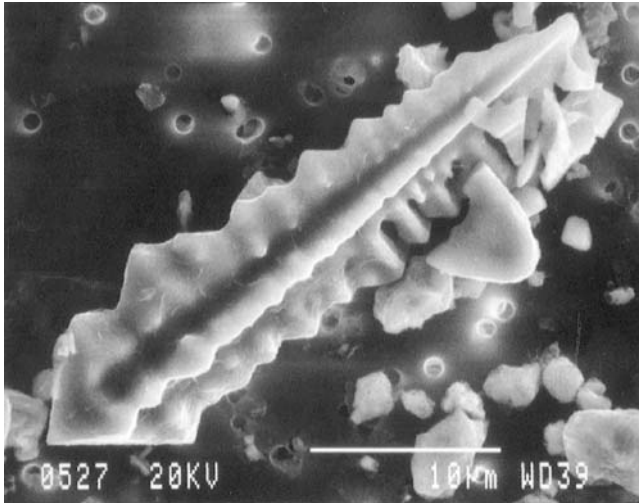


Fig. 9.17. Grain of Ni-rich spinel found in the KT boundary clay (photo courtesy of Eric Robin, CEA-CNRS)

cantly decreased the amount of light and energy received at the surface of the Earth.

9.3.3 Consequences for the Biosphere

The energy released by the Chicxulub impact was of about 10^{24} J. It remains now to evaluate how this enormous quantity of energy was transferred to the biosphere and led to the demise of organisms. The heat produced by the collision and the fallback of hot ejecta material must have induced forest fires over most of the continents. These fires released fine particles of soot (~ 50 Gt), traces of which are found in the KT clay (Wolbach and Anders 1989), and which contributed to darken further the Earth atmosphere (Melosh et al. 1990; Kring and Durda 2002). The vaporization of carbonate and evaporites from the target rock, indicates that along with the huge amount of dust released, the Chicxulub impact also injected vast quantities of carbon dioxide (350 to 3500 Gt), sulfur dioxide (40 to 560 Gt) and water vapor (200 Gt) in the upper atmosphere over a short period of time (for a review of the different models see Pierazzo et al. 2003). By blocking sunlight, the dust and the sulfate aerosols led to a break down of photosynthesis and an intense and brutal cooling of the Earth. Acid rain must have been produced by the conversion of sulfur dioxide to sulfuric acid, perhaps supplemented by the spontaneous reaction, under shock heating of the atmosphere, of oxygen and nitrogen to produce nitric acids, directly above the crater site (Lewis et al. 1982; Pope et al. 1997). The cooling period might have been followed by a longer period of greenhouse warming, as the carbon dioxide released by shock vaporization of the carbonate slowly

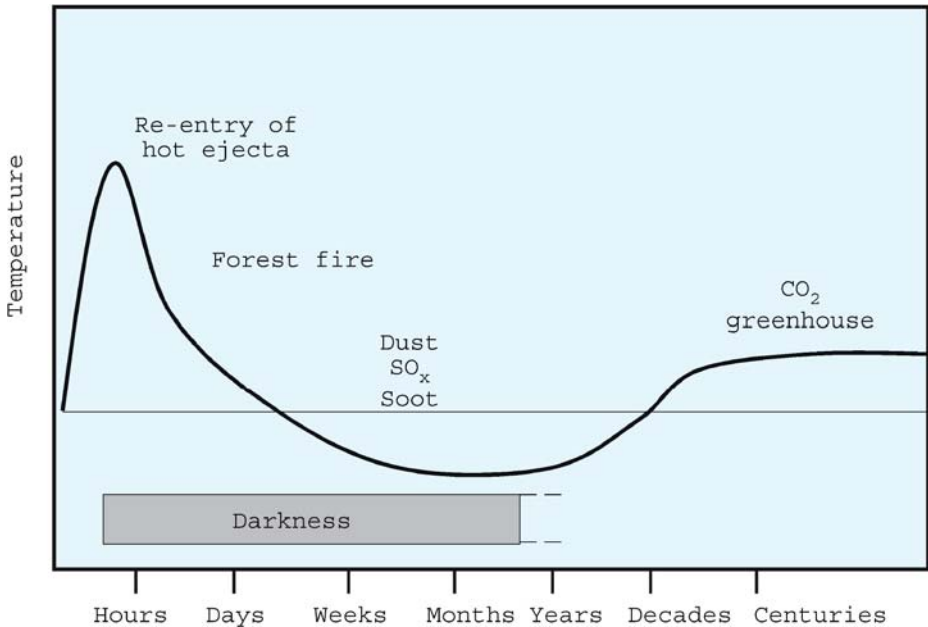


Fig. 9.18. Schematic representation of the perturbations affecting the Earth after the KT impact

accumulated in the atmosphere. These major perturbations (as well as others for example ozone hole formed by the bolide penetration) followed each other rapidly, leading to major and global climatic and oceanographic effects (Fig. 9.18).

Several authors have modeled the short and long time climatic consequences of the KT impact (see Pierazzo et al. 2003). The consequences on the world climate are at least two orders of magnitude superior to those of the eruption of the Pinatubo volcano in 1991 and they extended for several hundred to thousands of years after the impact. Models indicate global and rapid average-temperature changes between 5 and 15°C. Sea-surface water temperature fell by an estimated 3°C worldwide (Luder et al. 2002). The shock was radical for the Upper Cretaceous climate, which is considered much milder than today, with a lower temperature gradient between high and low latitude and much warmer oceans. The brutal cooling was followed by a major CO₂ warming. There remain many discussions concerning the exact magnitude and timing of the perturbations (Pope 2002). Stable isotope analyses (C and O) of both benthic and planktonic microorganisms just above the KT boundary indicate drastic temperature changes and a significant decrease of marine bioproductivity. Similar analyses also show that the biogeochemical cycles remained disturbed in the oceans for several hundred thousands of years after the impact (D'Hondt et al. 1998). The geographic distribution of dinoflagellates (unicellular marine organ-

isms) illustrates the oceanic crisis (Smit 1996). Sediments deposited in Tunisia show that directly after the KT impact boreal species migrated south, which can be interpreted as a sign of oceanic cooling. A few centimeters above in the same section, this trend is reversed as tropical dinoflagellates replace the boreal species (Galeotti et al. 2004). These micropaleontological observations clearly agree with the “cooling-warming” climatic scenario established by mathematical models.

9.3.4 Asteroid or Comet?

The origin of the projectile, asteroid or comet, should also be discussed. In theory, it has been suggested that the largest craters ($> 50\text{km}$) are essentially the result of cometary impacts (Shoemaker et al. 1990). However, the geochemical identification (PGE analyses) of the crater-forming projectiles does not seem to support this view (McDonald 2002; Tagle and Claeys 2004). Kyte (1998) extracted from KT boundary sediments in the Pacific Ocean, a small ($< 1\text{cm}$) fragment of carbonaceous chondrite, which could be a remnant of the Chicxulub projectile. The Cr isotopic signature of the KT boundary clay at different sites also substantiates a carbonaceous chondrite projectile (Shukolyukov and Lugmair 1998). This type of meteorite is relatively abundant in the asteroid belt; it appears thus likely that the Chicxulub projectile originated from the region between Mars and Jupiter.

9.4 Stratigraphic Distribution of Impact Events

9.4.1 In the Phanerozoic (0 to 540 Ma)

If impact events are to play a role in the evolution of the Earth, it is imperative to document their stratigraphic distribution. The number of known craters decreases after 50 million years and few craters older than 250Ma are identified on Earth (Fig. 9.2). The existing record is biased against older structures. This is expected considering the geological processes that tend to hide or erase older impact structures. The traces of these older events might be, as demonstrated for the KT boundary, preserved as ejecta particles in the sedimentary record. The ejecta debris currently identified are given in Fig. 9.19, and a very detailed discussion covering the last 250Ma is found in Montanari and Koeberl (2000).

The number of known ejecta deposits is inferior to that of known craters (compare Figs. 9.1 and 9.19). Craters and ejecta layers do not always correlate. For example, no crater has yet been identified to be associated with the most recent tektite strewnfield. It is perhaps located somewhere in the jungle of Indochina or buried under the active sedimentation of the Mekong River. The ejecta debris produced by the majority of craters have not been

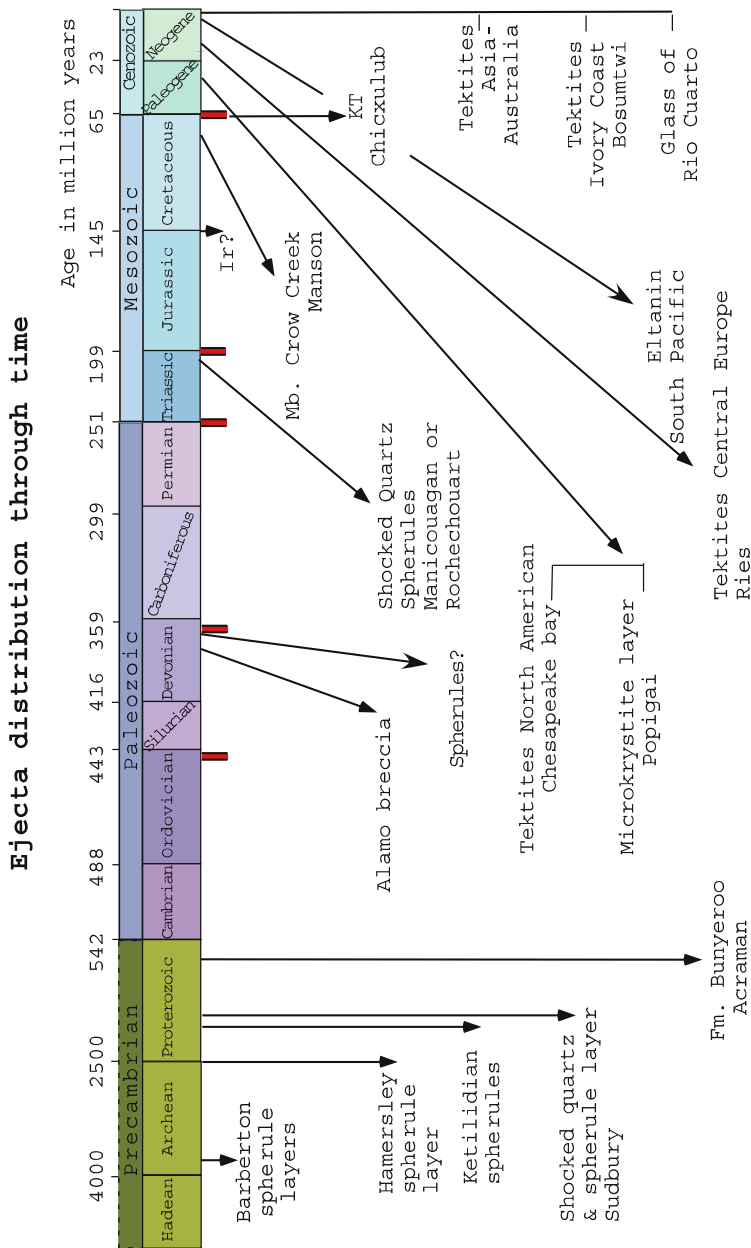


Fig. 9.19. Stratigraphic distribution of the ejecta layers known in the sedimentary record. A question mark (?) indicates levels where some controversy remains as to the identification of ejecta debris, the name of the source crater is indicated in italics when it is known, *F* = formation, *Mb* = member, the big five extinction horizons are marked in red

identified in the sedimentary record. This is in part due to the difficulty of precisely dating impact craters and/or to a lack of biostratigraphy resolution of the examined sections. Much research has also been devoted to linking impact events and stratigraphic boundaries marked by a major mass extinction. Despite a few discoveries, some promising others highly controversial, the KT boundary remains today the only stratigraphic horizon where a clear impact-extinction correlation can be established (see discussion in Alvarez 2003).

Figure 9.19, clearly shows that the number of ejecta layers identified in the Mesozoic and in particular in the Paleozoic is rather limited. Unless precisely constrained by biostratigraphy, it is particularly difficult to recognize a thin ejecta bed (a few mm to cm thick) in thick ancient sedimentary sequences because of weathering, diagenesis, metamorphism and tectonic activities. The very fine scale stratigraphic resolution and high precision correlations required pinpoint impact layers are often lacking in older sedimentary sequences. The analysis of the database developed for the KT boundary demonstrates that ejecta material tends to be better preserved in deep marine sequences (Kiessling and Claeys 2000; Claeys et al. 2002). Because of the almost complete subduction of the ancient ocean floor, these paleoenvironments are rarely preserved in the Paleozoic.

9.4.2 Proterozoic Impacts (540 Ma to 2.5 Ga)

Compared to the Phanerozoic, impact craters are rare in the Proterozoic. However, two of the largest impact structures on Earth, Sudbury in Canada and Vredefort in South Africa are dated at 1850 and 2023Ma, respectively. Their diameter is still subject to much discussion, 250km is estimated for Sudbury and Vredefort may even have reached 300km. Unfortunately, Sudbury is highly tectonized, and erosion has erased almost all but the deepest parts of the Vredefort structure. Sudbury is well-known to contain as part of its magmatic complex one of the richest Ni-Cu and PGE sulfide deposits in the world; its annual Ni production reaches 200 000T (Naldrett 2003). Both craters have been studied in detail (e.g., Grieve and Therriault 2000) and their formation must have, just as Chicxulub did, spread ejecta all over the world. Chadwick et al. (2001) had suggested that the impact spherules identified in the Ketilidian orogenesis in Greenland, which has an age range between 2130 and 1840Ma, could have originated from either of these events. However, considering the flux of object given in Table 9.1, other large craters may have formed during this ~300 million year period. Recently, a ~50cm-thick ejecta dated between 1878 and 1838Ma old was identified in several sections of the uppermost part of the Gunflint formation near the US-Canadian border in Minnesota and Ontario (Addison et al. 2005). This unit contains shocked quartz and impact spherules; it is most likely attributed to the nearby (650–800km to the east) Sudbury crater. A time gap of ~1.9Ga separates this layer from the oldest ejecta deposits identified in Australia (Simonson and Harnik 2000) and South Africa (Lowe et al. 2003).

9.4.3 Archean Impacts (2.5 to 4 Ga)

No crater is known in the Archean. Because of the intense level of erosion of ancient cratons, the preserved Archean sequences are often constituted of deep granitic lithologies. This does not ease the finding of impact craters, which, after all, are relatively superficial crustal structures. Here, impact evidence is essentially in the form of ejecta deposits. Together, the well-preserved and exposed sedimentary sequences of South Africa and of the Hamersley Basin in the west of Australia may contain up to 10 ejecta layers. They essentially consist of thick accumulations of impact spherules, whose size easily exceed 2 millimeters (Simonson and Harnik 2000). These spherules seem to be concentrated in two time windows, first between 2.64 and 2.49Ga for four Australian beds, and one layer found in the Transvaal Supergroup in South Africa. Some of those layers are probably correlated with each other and resulted from the same impact event. The South African Fig Tree and Overwacht Groups contains another four layers, dated between 3.22 and 3.46Ga. A last horizon, not precisely dated, may exist in the older Warrawoona Group of Australia. Although some controversy existed a few years ago, the impact origin of the majority of the layers is now established. Some are enriched in Ir and other PGEs, the spherules display morphologies quite similar to those found at the KT boundary, and Cr isotope systematically confirms the presence of an extraterrestrial component in the youngest horizon of the Fig Tree Group (Shukolyukov et al. 2000).

However, these layers present some differences with the ejecta recognized in the Phanerozoic. Simonson and Harnik (2000) estimate that the spherule concentration is greater than that measured in the more recent layers. The size of the spherules also seems to be somewhat larger and the ejecta layers are always thicker than one centimeter. Moreover, these layers rarely contain shocked quartz, which is one of the reasons why their impact origin was previously questioned. So far, shocked quartz grains are only identified in a sample of the Jeerinah layer in Australia, (Rasmussen and Koeberl 2004). Some of the Archean impact spherules also display a rather particular composition; the original glass is replaced by K-feldspar. It is thus impossible to estimate the primary glass composition. However, the replacement preserved the typical elongated and fibrous texture, which resembles that obtained by the rapid cooling of basaltic glasses. Phanerozoic spherules are most commonly replaced by smectite and their chemistry rather reflects the composition of the continental crust, richer in Si and Al. The mafic (rich in Fe and Mg) character of the Archean spherules, combined with the lack of shocked quartz may be interpreted as the result of impacts on the oceanic crust. This interpretation agrees with the possibility that before 2.5Ga the volume of oceanic crust was more important than the present-day 60% (McCulloch and Bennet 1994). It thus seems likely that collisions the size (or eventually larger) than the KT event took place during two distinct periods in the Archean (Lowe et al. 2003). According to these authors, some projectiles may have reached 20km in diameter. Is the present record coincidental to the exceptional preservation of two periods marked by a higher delivery rate of large

impactors, or do these layers provide indication that Archean was marked by a superior flux of extraterrestrial objects on Earth? This question might be answered by the on-going searches for ejecta deposits in other stratigraphic levels in the Archean.

9.4.4 Hadean Impacts (4.0 Ga to Formation of the Earth)

The oldest known terrestrial rocks are dated between 3.8 and 4.0 Ga, and the cores of a few detrital zircons provide a formation age dating back to 4.4 Ga (Martin 2001; Wilde et al. 2001; Valley et al. 2002). No traces remain, but even Sudbury and Vredefort would seem small compared to the craters that must have formed during the first 500 million years of the Earth. On other planets, whose surface was not remodeled by erosion, sedimentation and plate tectonics (Moon, Mars, Venus, etc.) the most ancient impact structures easily exceed 1000 km in diameter (see for example Stoeffler and Ryder 2001 for a recent revision). The battered surfaces of these planets witness past collisions which frequency and scale were two to perhaps three orders of magnitude higher than the Phanerozoic values. It seems highly unlikely that a target the size of Earth would have been spared, even if the record of these past impacts has been obliterated. Their signature is perhaps to be found in the existence of elevated concentrations of shocked minerals, tektites or PGE in sedimentary sequences at the very base of the Archean. So far, such searches have failed to yield convincing evidence (Koeberl et al. 2000). However, a possible meteoritic signal has been detected in the ~ 3.8 Ga-old metamorphosed sedimentary sequences of Isua (Greenland) based on the ^{182}Hf - ^{182}W isotopic ratios (Shoenberg et al. 2002). These results remain to be confirmed. A chromium isotope study carried out on similar lithologies at Isua failed to detect an extraterrestrial component (Frei and Rosing 2005).

During the Hadean, impacts were one of the dominant processes of Earth's evolution. The lack of lithological record between 4.5 and 3.9 Ga is in part explained by the late heavy bombardment (LHB) period between 4.0 and 3.9 Ga, that remodeled completely the surface of the Earth (Ryder et al. 2000; Kring and Cohen 2002). The existence of this bombardment is clearly established. However, the magnitude of the collisions between 4.5 and 3.9 Ga remains a topic of major discussions. Two explanations have been proposed (Fig. 9.20). The frequency of impacts declined very slowly and progressively since the end of the accretion period, which would be the reason why the bombardment is still important around 4.0 Ga. In this view, the so-called LHB is not seen as an exceptional event, but rather as the tailing of an era, lasting some 500 million years during which the Earth is subject to a series of major and devastating collisions, capable of repetitively melting the crust and vaporizing the oceans. Another view advocates a rapid decline of the frequency of impacts after the formation of the Moon to a value only slightly higher than today ($\sim 2X$). The period of between 4.0 and 3.9 represents thus an exceptional and cataclysmic event, the cause(s) of which remains to be explained. Today, the majority of authors favor the LHB

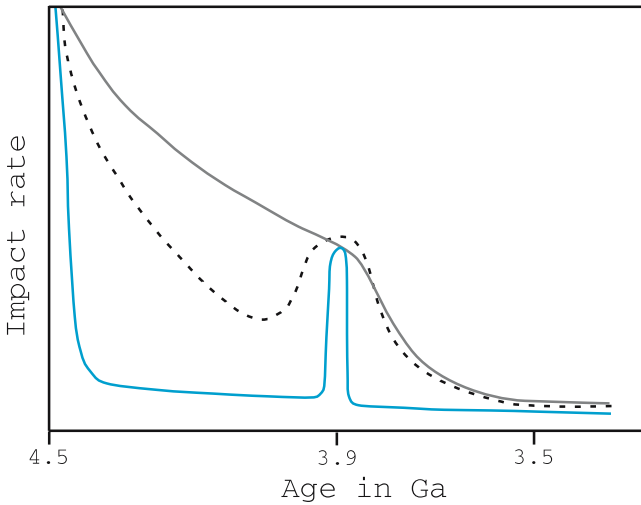


Fig. 9.20. Schematic representation of the flux of projectiles on Earth between 4.5 and 3.5 Ga (modified after Kring 2003). The *grey curve* shows a slow decrease of the bombardment, the *blue curve* illustrates the leading hypothesis of a rapid decrease of the flux followed by an unexplained peak, the late heavy bombardment, around 4.0–3.9 Ga. The duration of this high flux period varies (20 and 200 Ma); the slope of the decrease after 3.9 Ga is also poorly constrained. It cannot be excluded that other cataclysmic events took place between 4.5 and 4.0 Ga, but their traces must have been fully erased by the complete resetting of the Earth surface caused by the last event

cataclysmic scenario. It is supported by a series of arguments among them the fact that 500 million years of violent impacts should have left an obvious trace on the Moon. However, no impact melt rock older than 3.9 Ga is known (Stoeffler and Ryder 2001). On Earth, the oxygen isotopic signature ($\delta^{18}\text{O}$) of the oldest known zircons (dated at 4.4 Ga) indicates that they were formed under temperate surface conditions, compatible with the existence of liquid water, which seems to rule out an extended period of intense collisions (Valley et al. 2002).

On the Moon, where the record of these collisions is best preserved, some 1700 craters with a diameter > 20 km are known; among them 15 have diameter between 300 and 1200 km. All of them are dated between 4.0 and 3.9 Ga, ages that correspond to the Nectarian and the beginning of the Imbrian periods according to the Lunar stratigraphy (Wilhems 1987). On Earth, Kring (2003) estimates the formation of $\sim 22\,000$ craters with a diameter over 20 km based on an impact flux between 10 to 50 times larger than the one that hit the Moon. The number of structures larger than 1000 km would vary between 40 and 200, and it is not impossible that some would reach 5000 km (Grieve et al., 1990; Kring and Cohen 2002). The duration of the LHB cataclysmic event is estimated between 20 and 200 Ma, depending on the mass flux estimation used in the calculation. A collision capable of generating a 20 km impact crater could take place every

20 to 10 000 years (Kring 2003). A brief but important increase in the number of impact events followed by their decline marks this 4.0 to 3.9 Ga interval. Based on the lunar cratering record, it seems that around 3.8 Ga the impact frequency had decreased significantly and was stabilized more or less the present values. The dating of spherules formed during Lunar impacts using the ^{40}Ar - ^{39}Ar technique shows that this frequency decline is perhaps more progressive (by factor 2 or 3 between 3.5 Ga and 1 Ga), and that it has increased slightly again some 500 Ma ago (Culler et al. 2000). This interesting observation based on the dating of spherules collected by the Apollo 14 mission needs to be confirmed by applying the same approach to other sites sampled by Lunar missions.

9.5 Discussion: Impact, Origin of Life and Extinctions

The consequences for the biosphere of an impact event depend evidently of its magnitude and the total energy released. Nevertheless, other factors must also be considered. The lessons learned from the Chicxulub-KT boundary impact-extinction scenario might in some way help to understand the effects on the biosphere of large Precambrian impacts. Kring (2003) estimates that about 40 Chicxulub-size impacts may have occurred since 3.9 Ga, but their consequences for the climate and the biosphere are not directly comparable. The Chicxulub impact injected in the atmosphere large volumes of gases and dust capable of modifying the climate and the amount of sunlight received by the Earth, which ultimately led to the decline of the biosphere. The induced modifications of the global Earth system depend also of pre-existing environmental conditions, and of the *state* of the biological communities. For example, the injection of CO_2 in an atmosphere already rich in greenhouse gases does not produce the same climatic effects as in an atmosphere where such gases are in low concentrations (Kring 2003). Selsis and Parisot (2001) have discussed the evolution of the composition of the Earth atmosphere. During the Archean, the CO_2 contribution produced by the Chicxulub event would have represented $< 0.1\%$ of the total CO_2 already present in this early atmosphere. In all probability, it would have been unable to modify the existing climate system. Most likely, an impact induced greenhouse warming is only possible during the Phanerozoic (Kring 2003). In parallel, the period of darkness induced by the impact release of fine dust in the atmosphere only devastates photosynthetic organisms; it would have had no consequence for the Archaea and Bacteria communities of the Archean and the early Proterozoic. However, because of their way of life, and the different energy sources that these communities relied on, such organisms may have been more sensitive to other factors such as a cooling of water temperature and/or modification of the ocean chemical composition and biogeochemical cycles. Impact-generated acid rain would have had little effects on the organisms capable of living in the Archean oceans, whose pH was probably lower than 6 (Kasting and Toon 1989). The production of large forest fires is, of course, only possible after the evolution of terrestrial plants in the Silurian. The composition of the target rock and its

geographic distribution also play a role. Climate forcing gases such CO₂ and S-rich gases are released by the shock vaporization of carbonates and evaporites, yet these lithologies only become abundant respectively around 3.0Ga and in the middle of the Proterozoic. The outcome of an impact on oceanic or continental crust varies. It seems likely that during the Archean, most of the impacts hit oceanic crust (see Sect. 9.4.3). The effects of impact during the extended periods (> 10 million years) around 2400Ma and 700Ma during which the whole planet was fully glaciated (known as “snowball Earth” Hoffman et al. 1998; Kirschvink et al. 2000) must also be envisaged. This snowball Earth idea is today subject to vigorous debates. According to its supporters, the ice thickness reached several hundred meters above the oceans and several kilometers above the continents (Bodiselsitch et al. 2005). Little is known about the effects of impact events on icy targets (Dypvik et al. 2003), but it is possible that during such glaciated periods, they would lead to the production of major quantities of liquid water and/or would injected important volumes of H₂O vapor, also a greenhouse gas, into the atmosphere. These examples illustrate that one has to remain rather careful and open-minded when trying to simply transplant the Chicxulub-KT boundary scenario to the Precambrian. If during this period, impact events have affected the evolution of the biosphere, the mechanisms involved were most likely drastically different. In many ways, the existing environmental and biological parameters conditioned the consequences of impact events.

Before 3.9Ga, the possible effects of large impacts on the evolution of the early Earth and origin of life, remain poorly defined and are most commonly the result of models, according to rather vague constraints or even with a considerable dose of speculation. Megaimpacts could have formed basins where large volumes of water accumulated, perhaps forming sheltered sites favorable for the first biosyntheses or the development of life. They could have lead to the destruction or the fracturing of the crust, what might have contributed to the initiation or the intensification of some sort of plate tectonic process, or the rise of depth-rooted material to the surface. The production of massive volcanism by decompression melting of the lithosphere after an impact capable of perforating the oceanic crust may be considered here (Jones et al. 2002). According to these authors, such events might generate extensive basaltic lavas and formation of a large igneous province, which might also be accompanied by major outgassing. Impacts could also have affected the evolution of the atmosphere. A large collision could erode part of it (or all of it?) and new volatiles, some possibly coming from the projectile itself and/or from the vaporization of the target rock, could be incorporated in the newly formed atmosphere. Depending on their depths and morphology, total or partial vaporization of the oceans is also to be considered. Such vaporization would have had a sterilizing effect on the marine life or would have forced the early organisms to seek refuge in deep-water environments such as the hydrothermal vents located on the ocean floor (see Sleep et al. 1989; Zahnle and Sleep 1997). Less extreme consequences, such as severe changes in seawater temperature, may also have had drastic consequences for organisms

that inhabited the ancient oceans. If that was indeed the case, the early development of life could have followed a chaotic path between severe extinctions and origin of new forms. Eventually a contribution of extraterrestrial organic components imported by these projectiles should be considered. All these questions remain rather elusive and are based on a mixture of speculations and interesting hypotheses, which are based on the lack of record, and are particularly difficult to test or substantiate at this point.

Kring (2003) summarized with elegance the possible relationships between impacts and evolution of life during the Hadean. Four scenarios exist, with of course numerous ramifications:

1. The bombardment was too intense and devastating for life to originate or to evolve (Fig. 9.20).
2. Other factors than impact prevented its origin.
3. Life indeed started during the Hadean because the impact frequency was tolerable (Fig. 9.20) but it was completely extinct during the cataclysmic LHB around 4.0 to 3.9Ga. The whole process had to start from scratch, perhaps under another form, in the beginning of the Archean.
4. Life originated in the Hadean, evolved and diversified, which resulted in a rapid occupation of various milieus. At about 4.0 to 3.9Ga, the LHB led to the extinction of the vast majority of organisms. Only the thermophilic extremophiles living in the warm hydrothermal settings at the bottom of the deep oceans survived this cataclysmic period. When the impacts decreased in the early Archean, they evolved and diversified again to occupy the whole planet.

Is it possible that organic components present on the impacting projectile survived a hypervelocity impact? Meteorites, such as the carbonaceous chondrites, contain numerous amino acids and based on calculations comets would be even richer in these components (Chyba et al. 1990). The high temperatures generated during a collision event were believed to completely obliterate the complex organic molecules. This conclusion seems to remain valid for asteroid impacts. However, simulations now show that a proportion of amino acids is capable of surviving the impact of a comet in the ocean (Pierazzo and Chyba 1999). Comets contain ices, which are easily vaporized. The vapor cloud produced by such an impact is very energetic but cools off rather rapidly during its expansion in the atmosphere. The collision angle also plays an important role: an oblique impact induces lower temperatures in the vapor cloud, leading to a less complete decomposition of the organic molecules (Pierazzo and Chyba 1999). These authors made calculations that show that a comet impacting Earth at an angle lower than 15° could have dispersed about 10% of its amino acids into the early oceans. Such impacts are of course rather infrequent, but, considering the rates estimated for the Hadean and Early Archean, it is not impossible to imagine that an extraterrestrial contribution in amino acids or other organic molecules was delivered to the oceans. The key question is to know if this contribution was

significant enough to possibly influence in any way the origin and evolution of life.

Moreover, the large depressions created by impact events may perhaps represent settings favorable for the origin and/or evolution of life. In their centers, the heat generated by the large volume of melt rock induces an intense hydrothermal circulation (Ames et al., 1998; Naumov, 2002). Such impact crater hydrothermal systems would resemble in some ways their volcanic equivalents from the ocean floors that are often considered as having harbored the first living organisms (see for example Farmer, 2000). One major objection to life originating in volcanic hydrothermal vents is the high temperature of its fluids ($> 300^{\circ}\text{C}$) that would most certainly destroy any organic molecule immediately after its formation. The hydrothermal system of a crater would extend laterally with its ramifications covering most of the structure. In large Phanerozoic craters, hydrothermal circulation may persist for $\sim 1\text{Ma}$, as advocated by Abramov and Kring (2003) in the case of Chicxulub. The life span of a crater hydrothermal system appears thus sufficient to concentrate in its fluid a large number of complex organic molecules. Some contamination by the projectile might even be envisaged. The temperature of the circulating fluids decreases towards the periphery of the crater and with time as the impact melt pool crystallizes. Mineralogical studies at the Ries (25 km), Manson (35 km) and Puchez-Katunki (80 km) impact structures indicate temperature gradients ranging from 400° to less than 100°C (Newsom et al. 1986; Naumov 2002). Moreover, the highly fractured and brecciated lithologies on the flank of the crater would favor a very intense circulation of the fluids. In such environments, the fluids would also reach their lowest range of temperature. It is thus possible to speculate that the newly formed organic molecules would have been able to survive at temperatures significantly lower than those occurring in the black smoker chimneys on the ocean floor. Fractured zones, at the periphery of an impact crater containing an active hydrothermal system, meet some of the conditions favorable for the first biosyntheses and the early evolution of life, on Earth and, by extension, on other rocky planets, such as Mars.

References

- Abramov, O., and Kring, D.A. (2003). Finite-difference modeling of impact-induced hydrothermal systems. Lunar and Planetary Science Conference XXXIV, Abstract 1846. Lunar and Planetary Institute, Houston, TX.
- Addison, W.D., Brumpton, G.R., Vallini, D.A., McNaughton, N.J., Davis, D.W., Kissin, S.A., Fralick, P.W., and Hammond, A.L. (2005). Discovery of distal ejecta from the 1850 Ma Sudbury impact event. *Geology*, 33, 193–196.
- Ames, D.E., Watkinson, D.H., and Parrish, R.R. (1998). Dating of a hydrothermal system induced by the 1850 Ma Sudbury impact event. *Geology*, 26, 447–450.
- Ahrens, T.J., and O’Keefe, J.D. (1977). Equations of state and impact-induced shock-wave attenuation on the Moon. In: Roddy, D.J., Pepin, R.O., and Merrill, R.B.

- (eds.) *Impact and Explosion Cratering: Planetary and Terrestrial Implications*, pp. 639–656. Pergamon, New York.
- Alvarez, W. (1997). T. Rex and the Crater of Doom. Princeton University Press, Princeton, NJ.
- Alvarez, W. (2003). Comparing the evidence relevant to impact and flood basalt at times of major mass extinctions. *Astrobiology*, 3, 153–161.
- Alvarez, L.W., Alvarez, W., Asaro, F., and Michel, H.V. (1980). Extraterrestrial cause for the Cretaceous-Tertiary extinction. *Science*, 208, 1095–1108.
- Bottker, W.F.J., Nolan, M.C., Greenberg, R., and Kolvoord, R.A. (1994). Collisional lifetimes and impact statistics of near Earth asteroids. In: Gehrels, T. (ed.) *Hazards Due to Comets and Asteroids*, pp. 337–357. University of Arizona Press, Tucson, AZ.
- Bodisilitsch, B., Koeberl, C., Master, S., and Reimold, W.U. (2005). Estimating duration and intensity of neoproterozoic snowball glaciations from Ir anomalies. *Science*, 308, 239–242.
- Buffetaut, E. (1996). Tous les gros animaux disparaissent. *La Recherche*, 293, 65–67.
- Canup, R.M. (2004). Simulations of a late lunar-forming impact. *Icarus*, 168, 433–456.
- Chadwick, B., Claeys, P., and Simonson, B. (2001). New evidence for a large Palaeoproterozoic impact: spherules in a dolomite layer in the Ketilidian orogen. *Journal of the Geological Society of London*, 158, 331–340.
- Chyba, C.F., Thomas, P.J., Brookshaw, L. and Sagan, C. (1990). Cometary delivery of organic molecules to the early Earth. *Science*, 249, 366–373.
- Chyba, C.F., Owen, T.C., and Ip, W.H. (1994). Impact delivery of volatiles and organic molecules to Earth. In: Gehrels, T. (ed.) *Hazards Due to Comets and Asteroids*, pp. 9–58. University of Arizona Press, Tucson, AZ.
- Claeys, P. (1995). When the sky fell on our heads: identification and interpretation of impact products in the sedimentary record. *Reviews of Geophysics, Supplement*, U.S. National Report to International Union of Geodesy and Geophysics 1991–1994, pp. 95–100.
- Claeys, P., Kiessling, W., and Alvarez, W. (2002). Distribution of Chicxulub ejecta at the Cretaceous-Tertiary Boundary. In: Koeberl, C., MacLeod, K.G. (eds.) *Catastrophic Events and Mass Extinctions: Impacts and Beyond*, pp. 55–69. Geological Society of America Special Paper 356, Boulder, CO.
- Claeys, P., Heuschkel, S., Louvejeva-Baturina, E., Sanchez-Rubio, G., and Stöfler, D. (2003). The suevite in drill hole Yucatan 6 in the Chicxulub impact crater. *Meteoritics and Planetary Science*, 38, 1299–1317.
- Culler, T.S., Becker, T.A., Muller, R.A., and Renne, P.R. (2000). Lunar impact history from $^{40}\text{Ar}/^{39}\text{Ar}$ dating of glass spherules. *Science*, 287, 1785–1788.
- Deutsch, A., Grieve, R.A.F., Avermann, M., Bischoff, L., Brockmeyer, P., Buhl, D., Lakomy, R., Muller-Mohr, V., Ostermann, M., and Stöfler, D. (1995). The Sudbury structure (Ontario, Canada): a tectonically deformed multiring impact basin. *Geologische Rundschau*, 84, 697–709.
- D'Hondt, S., Donaghay, P., Zachos, J.C., Luttenberg, D., and Lindinger, M. (1998). Organic carbon flux and ecological recovery from the Cretaceous-Tertiary mass extinction. *Science*, 282, 276–278.
- Dietz, R.S. (1947). Meteorite impact suggested by the orientation of shatter-cones at the Kentland, Indiana disturbance. *Science*, 105, 42–43.

- Doukhan, J.C., and Leroux, H. (1996). La preuve par les quartz. *La Recherche*, 293, 56–57.
- Dypvik, H., Burchell, M.J., and Claeys, P. (2003). Impacts into marine and icy environments: a short review. In: Dypvik, H., Burchell, M.J., Claeys, P. (eds.) *Cratering in Marine Environments and on Ice*, pp. 1–19. Springer, Berlin Heidelberg New York.
- Dypvik, H., and Jansa, L.F. (2003). Sedimentary signatures and processes during marine bolide impacts: a review. *Sedimentary Geology*, 161, 309–337.
- Farmer, J.D. (2000). Hydrothermal systems: doorways to early biosphere evolution. *GSA Today*, 10, 1–8.
- Frei, R. and Rosing, M.T. (2005). Search for traces of the late heavy bombardment on Earth: results from high precision chromium isotopes. *Earth and Planetary Science Letters*, 236, 28–40.
- French, B.M. (1998). *Traces of catastrophe: a handbook of shock-metamorphic effects in terrestrial meteorite impact structures*. Lunar and Planetary Institute, Houston, TX.
- Galeotti, S., Brinkhuis, H., and Huber, M. (2004). Records of post-KT boundary millennial scale cooling from western Tethys: a smoking gun for the impact winter hypothesis? *Geology*, 32, 529–532.
- Gersonde, R., Deutsch, A., Ivanov, B.A., and Kyte, F.T. (2002). Oceanic impacts: a growing field of fundamental geosciences. *Deep-Sea Research II*, 49, 951–957.
- Goltrant, O., Cordier, P., and Doukhan, J.C. (1991). Planar deformation features in shocked quartz; a transmission electron microscopy investigation. *Earth and Planetary Science Letters*, 106, 103–115.
- Grieve, R.A.F., and Shoemaker, E.M. (1994). The record of past impacts on Earth. In: Gehrels, T. (ed.) *Hazards Due to Comets and Asteroids*, pp. 417–462. University of Arizona Press, Tucson, AZ.
- Grieve, R., and Therriault, A. (2000). Vredefort, Sudbury, Chicxulub: three of a kind? *Annual Review Earth Planetary Science*, 28, 305–338.
- Grieve, R.A.F., Pilkington, M., and Parmentier, E. (1990). Large impact basins and the early Earth, *International Workshop on Meteorite Impact on the Early Earth*, pp. 1819. Perth, Australia.
- Henkel, H., and Reimold, R.U. (1998). Integrated geophysical modelling of a giant, complex impact structure: anatomy of the Vredefort Structure, South Africa. *Tectonophysics*, 287, 1–20.
- Hildebrand, A.R., Penfield, G.T., Kring, D.A., Pilkington, M., Camargo Z.A., Jacobsen, S.B., and Boynton, W.V. (1991). Chicxulub crater: a possible Cretaceous/Tertiary boundary impact crater on the Yucatán Peninsula, Mexico. *Geology*, 19, 867–871.
- Hoffman, P.F., Kaufman, A.J., Halverson, G.P., and Schrag, D.P. (1998). A neoproterozoic snowball Earth. *Science*, 281, 51–54.
- Ingle, S., and Coffin, F.M. (2004). Impact origin for the greater Ontong Java Plateau? *Earth and Planetary Science Letters*, 218, 123–134.
- Ivanov, B.A., and Melosh, H.J. (2003). Impacts do not initiate volcanic eruptions. Eruption close to the crater. *Geology*, 31, 869–872.
- Jones, A.P., Claeys, P., and Heuschkel, S. (2000). Impact melting: a review of experimental constraints for carbonate targets and applications to the Chicxulub crater. In: Gilmour, I., and Koeberl, C. (eds.) *Impacts and the Early Earth*, pp. 343–362. *Lecture Notes in Earth Sciences*. Springer, Berlin Heidelberg New York.

- Jones, A.P., Price, G.D., Price, N.J., DeCarli, P.S., and Clegg, R.A. (2002). Impact induced melting and the development of large igneous provinces. *Earth and Planetary Science Letters*, 202, 551–561.
- Kasting, J.F., and Toon, O.B. (1989). Climate evolution of the terrestrial planets. In: Atreya, S.K., Pollack, J.B., Matthews, M.S. (eds.) *Origin and Evolution of Planetary and Satellite Atmospheres*, pp. 423–449. University of Arizona Press, Tucson, AZ.
- Kiessling, W., and Claeys, P. (2001). A geographic database approach to the K/T boundary. In: Buffetaut, E., Koeberl, C. (eds.) *Geological and Biological Effects of Impact Events*, pp. 83–140. Springer, Berlin Heidelberg New York.
- Kirschvink, J.L., Gaidos, E.J., Bertani, L.E., Beukes, N.J., Gutzmer, J., Maepa, L.N., and Steinberger, R.E. (2000). Paleoproterozoic snowball Earth: extreme climatic and geochemical global change and its biological consequences. *Proceedings National Academy of Science USA*, 97, 1400–1405.
- Koeberl, C. (1986). Geochemistry of tektites and impact glasses. *Annual Review of Earth Planetary Sciences*, 14, 323–350.
- Koeberl, C. (1992). Water content of glasses from the K/T boundary, Haiti: an indication of impact origin. *Geochimica et Cosmochimica Acta*, 56, 4312–4332.
- Koeberl, C. (1998). Identification of meteoritic component in impactites. In: Hutchinson, R., McCall, G.J.H., Rothery, R.A. (eds.) *Meteorites: Flux with Time and Impact Effects*, pp. 133–153. Geological Society Special Publications 140, London.
- Koeberl, C. (2001). Craters on the moon from Galileo to Wegener: a short history of the impact hypothesis, and implications for the study of terrestrial impact craters. *Earth, Moon and Planets*, 85–86, 209–224.
- Koeberl, C., Reimold, W.U., and Shirey, S.B. (1998). The Aouelloul crater, Mauritania: on the problem of confirming the impact origin of a small crater. *Meteoritics and Planetary Science*, 33, 513–517.
- Koeberl, C., Reimold, W.U., McDonald, I., and Rosing, M. (2000). Search for petrographic and geochemical evidence for the late heavy bombardment on Earth in Early Archaean rocks from Isua, Greenland. In: Gilmour, I. and Koeberl, C. (eds.) *Impacts and the Early Earth*, pp. 73–97. Springer, Berlin Heidelberg New York.
- Koeberl C., Peucker-Ehrenbrink B., Reimold W.U., Shukolyukov A., and Lugmair G.W. (2002). Comparison of Os and Cr isotopic methods for the detection of meteoritic components in impactites: examples from the Morokweng and Vredefort impact structures, South Africa. In: Koeberl, C. and MacLeod, K.G. (eds.) *Catastrophic Events and Mass Extinctions: Impacts and Beyond*, vol. 356, pp. 607–617. Geological Society of America Special Paper, Boulder, CO.
- Kring, D.A. (2003). Environmental consequences of impact cratering events as a function of ambient conditions on Earth. *Astrobiology*, 3, 133–152.
- Kring, D.A., and Cohen, B.A. (2002). Cataclysmic bombardment throughout the inner solar system 3.9–4.0 Ga. *Journal of Geophysical Research*, 107(E2), 5009.
- Kring, D.A., and Durda, D.D. (2002). Trajectories and distribution of material ejected from the Chicxulub impact crater: implications for post-impact wildfires. *Journal of Geophysical Research*, 107(E8), 5062.
- Kyte, F.T. (1998). A meteorite from the Cretaceous/Tertiary boundary. *Nature*, 396, 237–239.

- Kyte, F.T., and Bostwick, J.A. (1995). Magnesioferrite spinel in Cretaceous-Tertiary boundary sediments of the Pacific basin: Remnants of hot early ejecta from the Chicxulub impact? *Earth and Planetary Science Letters*, 132, 113–127.
- Kyte, F.T., Zhou, Z., and Wasson, J. (1981). High noble metal concentrations in a late Pliocene sediment. *Nature*, 292, 417–420.
- Lewis, J.S., Watkins, G.H., Hartman, H., and Prinn, R.G. (1982). Chemical consequences of major impact events on Earth. In: Silver, L.T., and Schultz, P.H. (eds.) *Geological Implications of Impacts of Large Asteroids and Comets on Earth*, pp. 215–221. Geological Society of America Special Paper 190, Boulder, CO.
- López Ramos, E. (1975). Geological summary of the Yucatán Peninsula. In: Nairn, A.E.M., and Stehli, F.G. (eds.) *The Ocean Basins and Margins*, vol 3: the Gulf of Mexico and the Caribbean, pp. 257–282. Plenum, New York.
- Lowe, D.R., Byerly, G.R., Kyte, F.T., Shukolyukov, A., Asaro, F., and 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.
- Luder, T., Benz, W., and Stocker, T.F. (2002). Modeling long-term climatic effects of impacts: first results. In: Nairn, A.E.M., and Stehli, F.G. (eds.) *Catastrophic Events and Mass Extinctions: Impacts and Beyond*, pp. 717–729. Geological Society of America Special Paper 356, Boulder, CO.
- Lyons, J.B., Officer, C.B., Borella, P.E., and Lahodinsky, R. (1993). Planar lamellar substructures in quartz. *Earth and Planetary Science Letters*, 119, 431–440.
- Margolis, S.V., Claey, P., and Kyte, F.T. (1991). Microtektites, microkrystites, and spinels from a late Pliocene asteroid impact in the Southern ocean. *Science*, 251, 1594–1597.
- Martin, H. (2001). Evolution géologique et géochimique de la terre: Genèse et évolution de la croûte primitive. In: Gargaud, M., Despois, D. and Parisot, J.-P. (eds.) *L'environnement de la terre primitive*, pp. 263–286. Presses Universitaires de Bordeaux, Bordeaux.
- McCulloch, M.T., and Bennet, V.C. (1994). Progressive growth of the Earth's continental crust and depleted mantle: geochemical constraints. *Geochimica et Cosmochimica Acta*, 58, 4717–4738.
- McDonald, I. (2002). Clearwater East impact structure: A re-interpretation of the projectile type using new platinum-group element data. *Meteoritics and Planetary Science*, 37, 459–464.
- Melosh, H.J. (1989). *Impact Cratering: a Geologic Process*. Oxford University Press, New York.
- Melosh, H.J., Schneider, N.M., Zahnle, K.J., and Latham, D. (1990). Ignition of global wildfires at the Cretaceous-Tertiary Boundary. *Nature*, 343, 251–254.
- Montanari, A., and Koeberl, C. (2000). *Impact Stratigraphy*. Springer, Berlin Heidelberg New York.
- Morgan, J., Warner, M., Brittan, J., Buffler, R., Camargo, A., Christeson, G., Dentons, P., Hildebrand, A., Hobbs, R., MacIntyre, H., Mackenzie, G., Maguire, P., Marin, L., Nakamura, Y., Pilkington, M., Sharpton, V., and Snyders, D. (1997). Size and morphology of the Chicxulub impact crater. *Nature*, 390, 472–476.
- Morgan, J.V., Warner, M.R., and Grieve, R.A.F. (2002). Geophysical constraints on the size and structure of the Chicxulub impact crater. In: Koeberl, C. and MacLeod,

- K.G. (eds.) *Catastrophic Events and Mass Extinctions: Impacts and Beyond*, pp. 39–46. Geological Society of America Special Paper 356, Boulder, CO.
- Naldrett, A.J. (2003). From impact to riches: evolution of geological understanding as seen at Sudbury, Canada. *GSA Today*, 13, 4–9.
- Naumov, M. (2002). Impact-generated hydrothermal systems: data from Popigai, Kara, and Puchezh-Katunki impact structures. In: Plado, J. and Pesonen, L.J. (eds.) *Impacts in Precambrian Shields*, pp. 117–172. Springer, Berlin Heidelberg New York.
- Newsom, H.E., Graup, G., Sowards, T., and Keil, K. (1986). Fluidization and hydrothermal alteration of the suevite deposit at the Ries crater, West Germany, and implications for Mars. *Journal of Geophysical Research*, 91, E239–E251.
- Neukum, G., and Ivanov, B.A. (1994). Crater size distributions and impact probabilities on Earth from lunar terrestrial planets and asteroid data cratering data. In: Gehrels, T. (ed.) *Hazards Due to Comets and Asteroids*, pp. 359–416. University of Arizona Press, Tucson, AZ.
- Penfield, G.T., and Camargo Z.A. (1981). Definition of a major igneous zone in the central Yucatán platform with aeromagnetism and gravity. *Society of Exploration Geophysicists Technical Program, Abstracts, and Biographies* 51, 37.
- Pesonen, L.J. (1996). The impact cratering record of Fennoscandia. *Earth, Moon and Planets*, 72, 377–393.
- Pesonen, L.J., and Henkel, H. (1992). Terrestrial impact craters and craterform structures, with a special focus on Fennoscandia. *Tectonophysics*, 216, 377–393.
- Pierazzo, E., Kring, D.A., and Melosh, J.H. (1998). Hydrocode simulation of the Chicxulub impact event and the production of climatically active gases. *Journal of Geophysical Research*, 103, 28607–28625.
- Pierazzo, E., and Chyba, C.F. (1999). Amino acid survival in large cometary impacts. *Meteoritics and Planetary Science*, 34, 909–918.
- Pierazzo, E., Hahmann, A., and Sloan, L.C. (2003). Chicxulub and climate: radiative perturbations of impact-produced S-bearing gases. *Astrobiology*, 3, 99–118.
- Pilkington, M., and Grieve, R.A.F. (1992). The geophysical signature of impact craters. *Reviews of Geophysics*, 30, 161–181.
- Pilkington, M., Hildebrand, A.R., and Ortiz-Aleman, C. (1994). Gravity and magnetic field modeling and structure of the Chicxulub Crater, Mexico. *Journal of Geophysical Research*, 99, 13,147–13,162.
- Pope, K.O. (2002). Impact dust not the cause of the Cretaceous-Tertiary mass extinction. *Geology*, 30, 99–102.
- Pope, K.O., Baines, K.H., Ocampo, A.C., and Ivanov, B. (1997). Energy, volatile production, and climate effects of the Chicxulub Cretaceous/Tertiary boundary impact. *Journal of Geophysical Research*, 102, 21,645–21,664.
- O’Keefe, J.D., and Ahrens, T.J. (1977). Impact induced energy partitioning, melting, and vaporization on terrestrial planets. *Proceedings of the 8th Lunar Science Conference*, pp. 3357–3374. Pergamon Press, New York.
- O’Keefe, J.D., and Ahrens, T.J. (1991). Tsunamis from giant impacts on solid planets, *Lunar Planetary Science Conference XXI*, 997, Lunar and Planetary Institute, Houston, TX.
- O’Keefe, J.D., and Ahrens, T.J. (1993). Planetary cratering mechanics. *Journal of Geophysical Research*, 87, A457–A463.

- Renne, P.R., Swisher, C.C., Deino, A.L., Karnet, D.B., and Owens, T.L. (1998). Intercalibration of standards absolute ages and uncertainties in $^{40}\text{Ar}/^{39}\text{Ar}$ dating. *Chemical Geology*, 145, 117–152.
- Robin, E., Bonté, P., Froget, L., Jéhanno, C., and Rocchia, R. (1992). Formation of spinels in cosmic objects during atmospheric entry: a clue to the Cretaceous-Tertiary boundary event. *Earth and Planetary Science Letters*, 108, 181–190.
- Robin, E. (1996). Le verdict du spinelle. *La Recherche*, 293, pp. 58–60.
- Rocchia, R. (1996). Naissance d'une théorie, d'une anomalie en iridium à la catastrophe cosmique. *La Recherche*, 293, 52–55.
- Roddy, D.J., Pepin, R.O., and Merrill, R.B. (1977). *Impact and explosion cratering: planetary and terrestrial implications*. Pergamon Press, New York.
- Rondot, J. (2003). *Les impacts météoritiques: Charlevoix et le Ries*. 183 p. MNH, Chabanel, Quebec.
- Ryder G., Koeberl, C., and Mojzsis, S.J. (2000). Heavy bombardment on the Earth at ~ 3.85 : the search for petrographical and geochemical evidence. In: Canup, R.M., and Richter, K. (eds.) *Origin of the Earth and Moon*, pp. 475–492. University of Arizona Press, Tucson, AZ.
- Selsis, F., and Parisot, J.-P. (2001). L'atmosphère primitive de la terre et son évolution. In: Gargaud, M., Despois, D., and Parisot, J.-P. (eds.) *L'environnement de la terre primitive*, pp. 216–233. Presses Universitaires de Bordeaux, Bordeaux.
- Sharpton, V.L., Dalrymple, G.B., Marin, L.E., Ryder, G., Schuraytz, B.C., and Urrutia, J. (1992). New links between the Chicxulub impact structure and the Cretaceous/Tertiary boundary. *Nature*, 359, 819–821.
- Shoemaker, E.M., Wolfe, R.F., and Shoemaker, C.S. (1990). Asteroid and comet flux in the neighborhood of Earth. In: Sharpton, V.L., and Ward, P.D. (eds.) *Global Catastrophes in Earth History*, pp. 155–170. Geological Society of America, Special Paper 247, Denver, CO.
- Schoenberg, R., Kamber, B.S., Collerson, K.D., and Moorbath, S. (2002). Tungsten isotope evidence from ~ 3.8 -Gyr metamorphosed sediments for early meteorite bombardment of the Earth. *Nature*, 418, 403–405.
- Shukolyukov, A., and Lugmair, G.W. (1998). Isotopic evidence for the Cretaceous-Tertiary impactor and its type. *Science*, 282, 927–929.
- Shukolyukov, A., Kyte, F.T., Lugmair, G.W., Lowe, D.R., and Byerly, G.R. (2000). The oldest impact deposits on Earth: first confirmation of an extraterrestrial component. In: Gilmour, I., and Koeberl, C. (eds.) *Impacts and the Early Earth*, pp. 99–115. Springer, Berlin Heidelberg New York.
- Simonson, B.M., and Harnik, P. (2000). Have distal ejecta changed through geologic time? *Geology*, 28, 975–978.
- Sleep, N.H., Zahnle, K.J., Kasting, J.F., and Morowitz, H.J. (1989). Annihilation of ecosystems by large asteroid impacts on the early Earth. *Nature*, 342, 139–142.
- Smit, J. (1996). Un épisode tragique: L'océan Folamour. *La Recherche*, 293, 62–64.
- Smit, J. (1999). The global stratigraphy of the Cretaceous-Tertiary boundary impact ejecta. *Annual Review of Earth and Planetary Sciences*, 27, 75–113.
- Smit, J., and Hertogen, J. (1980). An extraterrestrial event at the Cretaceous-Tertiary boundary. *Nature*, 285, 198–200.
- Spudis, P.D. (1993). *The Geology of Multiring Impact Basins: the Moon and Other Planets*. Cambridge University Press, New York.

- Stoeffler, D., and Langenhorst, F. (1994). Shock metamorphism of quartz in nature and experiment: 1. Basic observation and theory. *Meteoritics*, 29, 155–181.
- Stoeffler, D., and Ryder, G. (2001). Stratigraphy and isotopic ages of lunar geologic units: chronological standard for the inner solar system. In: *The Evolution of Mars*. 7-53. Space Science Reviews. International Space Science Institute, Bern, Switzerland.
- Swisher, C.C., Grajales-Nishimura, J.M., Montanari, A., Margolis, S.V., Claeys, P., Alvarez, W., Renne, P., Cedillo-Pardo, E., Maurrasse, F., Curtis, G.H., Smit J., and McWilliams, M.O. (1992). Coeval $^{40}\text{Ar}/^{39}\text{Ar}$ ages of 65.0 million years ago from Chicxulub crater melt rock and Cretaceous-Tertiary boundary tektites. *Science*, 257, 954–958.
- Tagle, R., and Claeys, P. (2004). Comet or asteroid shower in the late Eocene? *Science*, 305, 492–492.
- Tsikalas, F., Gudlaugsson, S.T., Eldholm, O., and Faleide, J.I. (1998). Integrated geophysical analysis supporting the impact origin of the Mjølnir structure, Barents Sea. *Tectonophysics*, 289, 257–280.
- Urrutia-Fucugauchi, J., Morgan, J., Stoeffler, D., and Claeys, P. (2004). The Chicxulub scientific drilling project (CSDP). *Meteoritics and Planetary Science*, 39, 787–790.
- Valley, J.W., Peck, W.H., King, E.M., and Wilde, S.A. (2002). A cool early Earth. *Geology*, 30, 351–354.
- Wilde, S.A., Valley J.W., Peck W.H., Graham, C.M. (2001). Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature*, 409, 175–181.
- Wilhems, D.E. (1987). *Geologic history of the Moon*. US Geological Survey, Professional Paper 1348. 302 p. Reston, Virginia.
- Wolbach, W.S., and Anders, E. (1989). Elemental carbon in sediments: determination and isotopic analysis in the presence of kerogen. *Geochimica et Cosmochimica Acta*, 53, 1637–1647.
- Zahnle, K.J., and Sleep, N.H. (1997). Impacts and the early evolution of life. In: Thomas, P.J., Chyba, C.F., and McKay, C.P. (eds.) *Comets and the Origin and Evolution of Life*, pp. 175–208. Springer, Berlin Heidelberg New York.

10 Towards a Global Earth Regulation

Philippe Bertrand

The regulation of a system is the way by which this system is maintained in a steady state, generally far from its thermodynamic equilibrium. One necessary, although not sufficient, condition for a system to be regarded as living is to be open and regulated. To maintain themselves, the living systems need to reduce their entropy by degrading energy from outside. Irreversible building processes such as biosynthesis and reproduction act permanently to maintain the living systems far from the thermodynamic equilibrium. However, any steady state is absolute or permanent. Such a concept is basically statistic, making sense only within our common perception field, far from the time and space scales where laws of quantum or general relativity theories apply. From another side, even if the term of ecological equilibrium is often used as a reference for our anthropically impacted environment, the species, populations, communities and their physical environments evolve slowly but inexorably. Therefore, the concept of regulation used in this paper is a simplification to express the stability of a steady state of a system within a given period. It can also be described as the ability of an organism, a population, or an ecosystem to survive external perturbations, which likened to its health robustness. In humans, plants, or bacteria, the building and repairing processes are unavoidably interrupted by ageing and death. However, life acts in this case like an advised businessman. It pays its taxes but, through reproduction, has generally already transferred and improved its capital towards investments into following and evolving generations. This way, the regulation, which is interrupted at the scale of an individual living system, may be preserved at the scale of a living system of higher order such as a population, a community, or a more or less extended ecosystem.

The life on the present Earth results from such investments that have been renewed for more than 3.5Ga through countless regulating processes. Some of the physical features of the present Earth, such as the composition of the atmosphere or oceans, show that the influence of regulating mechanisms is global and that, therefore, the whole Earth system should be considered as an ecosystem. What are the regulating mechanisms at the global scale? How old are they? Is the present Earth regulated and how robust is this regulation (Stengers 2003)? In the scientific context of exobiology, besides questions related to the environment-dependent emerging life on the primitive Earth, those of the emergence of a “Gaïan world” where living entities, their evolution (Kupiec and

Sonigo 2000) and their physical environment have become interdependent (homeostasis), must also be addressed.

As illustrated by the simple theoretical model “Daisy World” (Lovelock 1990), such an interdependence increases the duration of life-compatible conditions on a Earth-like planet as well as the robustness of the life regulation. It takes place without any predetermined intention of the system (teleology), which allows to exclude any mystic cause from the explanation of the observed phenomenon or effects. For about twenty years, biogeochemistry, a new research field, whose objective is the study of the interactions between the living systems and their physical environment at any spatial and timescales, has made considerable advances in describing the Earth’s regulating processes. Based on some examples, this paper gives a synthetic vision of this knowledge and highlights the more significant implications, some of them being still potential, working hypotheses. While the main biogeochemical cycles of oxygen, nitrogen, phosphorous and carbon are successively presented in this article, one of their main and common features is that they have strongly interacted to lead the Earth system towards an increasingly regulated situation, in spite of perturbations induced by either external (insolation, asteroids) or internal (tectonic, volcanism) forcing. This question is mainly addressed and discussed in the last section of this article.

10.1 The Oxygen: an Energy Story

Schlesinger said that the release of O_2 to an anaerobic Earth is perhaps the strongest reminder we have for the influence of biota on the geochemistry of the Earth’s surface (1991).

The high O_2 content of the present atmosphere (21% vol.) is one of the main features by which the Earth is distinguished from comparable rocky planets where life is absent, such as Mars and Venus. Moreover, this O_2 content results from permanent exchanges between ocean, atmosphere and continental surface, which means that most of the Earth ecosystem is submitted to the abundance and availability of molecular oxygen. The organisms living in this “oxysphere” result from an adaptive evolution and use preferably O_2 for their respiratory metabolism. Even anaerobic bacterial activities which survive today in some confined environments (guts, anoxic sediments, anoxic parts of stratified water columns, etc.) are indirectly based on O_2 . Indeed, recycling the anaerobic oxidants (nitrate, sulfate, Fe or Mn oxides, and organic-bond oxygen) depends on both the O_2 availability at the redox fronts (Fig. 10.1) and the organic matter produced by biosyntheses in oxic environments.

The very strong chemical reactivity of O_2 makes it extremely unstable. Maintaining the present-day availability of O_2 in the global ecosystem implies therefore that oxygen consumption must be fully compensated by an equivalent O_2 production through oxygenic photosynthesis. The simplified sto-

stored by photosynthesis. Therefore, when speaking about the global history and regulation of oxygen, we also refer to the slow transformation of the energy budget (or metabolism) of the global ecosystem towards its present day situation.

The O₂ content of the atmosphere is thought to have progressively increased since the early Proterozoic, about 2.3 Ga ago, with a faster step between 1 Ga and some 540 Ma ago (Kah et al. 2004). After this step, the O₂ atmospheric content would have reached its present level at the beginning of continental life development (Early Phanerozoic, Fig. 10.2). During the Phanerozoic, this content would have fluctuated around 21% vol., corresponding to a partial pressure of 21 kPa in the present atmosphere (Berner and Canfield 1989; Berner 2004). Fluctuations would have been between 15 and 40 kPa. Such estimations, based on paleontological observations or indirect geochemical measurements, are still subject to large uncertainties but seem to indicate that an oxygen regulation has prevailed for the last 500 millions years. Indeed, an atmospheric O₂ content lower than 15% is incompatible with big forest or vegetation fires, whose past occurrence is indicated by the massive deposition of coaly particles (pyrofusinite) within lacustrine sediment layers (Watson et al. 1978; Cope et al. 1985). On the other side, O₂ partial pressure higher than 21 kPa can favor, up to a given threshold, the living organisms that are able to adjust their respiratory flux in order to distribute O₂ to their cells at an optimized partial pressure, between 1 and 3 kPa (Massabuau 2003). Above 3 kPa, O₂ partial pressure begins to cause lethal lesions in eukaryote cells (personal communication with Massabuau, 2005). Although oxygen offers some energetic advantages, it is an extremely toxic el-

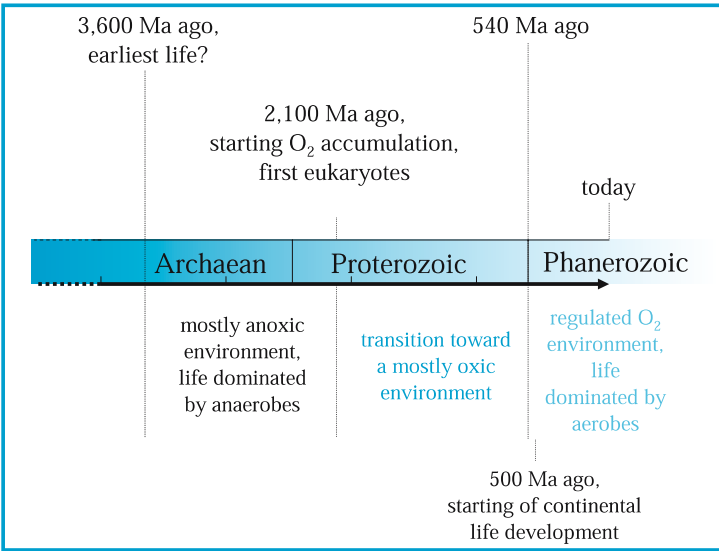


Fig. 10.2. Chronology of the main steps of the global Earth's ecosystem evolution

ement, the main responsible of our ageing-related diseases. The occurrence of a permocarboniferous episode of O_2 partial pressure much higher than today (35%) was proposed by Berner (2004). A large part of species survived this episode, which, because of constraints due to breathing physiology, would indicate that the O_2 stress for the biota was moderate in term of either intensity or duration. Indeed, a diver breathing an O_2 -enriched gas undergoes important lesions, which become mortal after 48 hours. However, other than these episodes, the Phanerozoic is mainly characterized by an O_2 regulation system (Lenton and Watson 2000b). By contrast, previous periods when life was present on Earth (Fig. 10.2) were characterized by anoxia (Archaean, Early Proterozoic) or raising O_2 concentration up to the present level, including probable large fluctuations (Late Proterozoic). What was the timing of this evolution and of the related processes?

On the present Earth, most of photosynthesis-produced oxygen is used either directly or indirectly by respiratory metabolisms to re-mineralize organic matter. This has probably not always been true. The primitive photosynthesis has probably given rise to the preservation of a significant part of the produced organic matter because most of the global environment was anoxic and because O_2 -consumers were not yet present or constituted a minor part of the biota. Besides, even if organic matter may have been re-mineralized by primitive anaerobes, this degradation should have favored the sedimentary fixation of reduced chemical species. For example because of the lack of O_2 , H_2S produced by sulfate bacteria (Fig. 10.1) probably could not be recycled into sulfate, and, consequently, may have been partly fixed in sediments as iron sulfides or organic sulfur compounds. Such conditions should have favored a large excess of the photosynthesis/respiration long-term budget, resulting in an increase of organic carbon and sulfide sedimentary storage and a related O_2 net production. Since the onset of life on Earth, most of the sedimentary rocks have been recycled several times by tectonic activity. Today, sedimentary carbon and sulfur occur mainly as carbonates and sulfates, respectively. One hypothesis that will be discussed later is that this situation results from a progressive transformation of the sedimentary reservoir, allowed by the geological recycling, and focused by the physicochemistry of the ocean.

For hundreds of millions of years, produced O_2 , because of its high chemical reactivity, was used to oxidize chemical species (elemental S, sulfides, Fe^{2+} , etc.) that have been present in ocean and atmosphere, or in contact with them. Occurrence of Precambrian Fe-enriched sediments (BIF = Banded Iron Formations), as well as the increase of the sulfate reservoir, are good indicators for this old biogeochemical activity. Since about 2.1 Ga, O_2 has been present at low partial pressures (around some kPa), but sufficient to allow the evolved development of aerobes. This development may also have been favored locally, around O_2 -production pools, similar to the present day high productive area in the ocean. This episode also corresponds with the development of the first eukaryotes. After the endosymbiosis theory (Margulis and Sagan 1989), the association of differ-

ent primitive prokaryotes, including the precursors of the mitochondria, would have been a key step for this evolution towards aerobes. Indeed, in the new endosymbiotic cells (eukaryotes), mitochondria can use O_2 and ATP to release the previously stored energy. At the end of the Proterozoic, some 540 Ma ago, exhaustion of the “chemical preys” for O_2 (the accessible reduced compounds or elements that can be oxidized by O_2) led to the increase of both the ocean and atmosphere O_2 reservoirs. Then, during the Phanerozoic, these reservoirs were probably regulated around their present levels. Due to the increase of atmospheric O_2 at the Proterozoic-Phanerozoic transition, and to the subsequent building of a stratospheric O_3 layer, UV radiations were reduced on the continental surfaces, which allowed the continental aerobes to develop. The O_2 -based respiration, either marine or continental, has been a key process in regulating O_2 . Without it, continuously increasing O_2 content may have transformed the

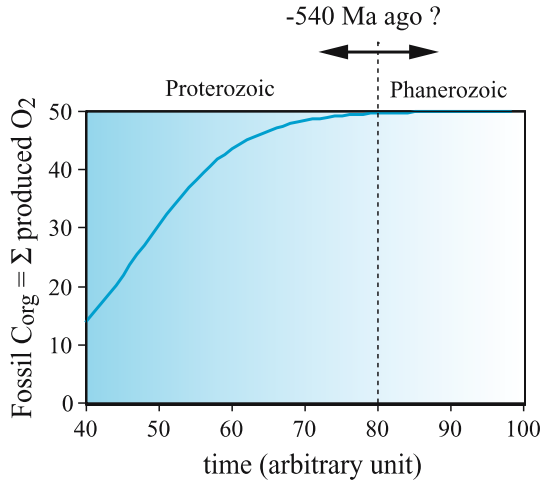


Fig. 10.3. The sedimentary organic matter accumulation implies that the corresponding amount of O_2 produced through photosynthesis is not consumed by respiration. On the other hand, the increasing abundance of O_2 , as a highly reactive compound, is unfavorable to the organic matter fossilization. Beyond O_2 consumption by reduced inorganic chemical species, a global regulation, implying an equivalent rate of sedimentary organic matter fossilization and respiration, has progressively settled. The final step of this evolution may be schematized with a simple function where the growth is limited by an abundance level of one of its products. Since the beginning of the Phanerozoic (540 Ma ago), the abundance of accumulated O_2 in the environment together with the evolving level of aerobes, have been probably sufficient to respire as much of organic matter as photosynthetically produced. Consequently, the growth of O_2 abundance in the global environment has probably been stopped. This may explain why the atmospheric O_2 partial pressure, which is dependent on this abundance, is thought to have been broadly stabilized—in spite of some temporary perturbations—during the Phanerozoic

whole planet in a deadly toxic trap. On this new aerobic planet, the global photosynthesis/respiration budget may be considered as balanced on the long-term, which means that O_2 production equals O_2 consumption (Fig. 10.3). This balance implies that there is no or few net accumulation of sedimentary organic matter since the Early Phanerozoic. This is a counter-intuitive idea for us, as fossil fuel consumers. However, successive oceanographic and geological settings may have favored locally and temporarily the formation of fossil fuels (Bertrand and Lallier-Vergès 1993; Stow et al. 2001), while the re-mineralization of previously accumulated organic matter by aerobes may have dominated in other places.

10.2 Nitrogen and Phosphorus: the Nutrient Feedback

In the present ocean, the dissolved nitrogen (N)/phosphorous (P) atomic ratio (16) is surprisingly similar to the biological demand as assessed by the phytoplankton biomass composition (Redfield 1958). Several recent studies (for example Lenton and Watson 2000a) have shown that such similarity does not simply result from the adjustment of the biological demand to the constraint of its chemical environment. This pointed out that they rather result from a more sophisticated regulation coming from both the evolution of the biological producers and their feedback on the chemical ocean environment, thus enhancing the reduction of the N/P contrast between the biological demand and the dissolved nutrient availability. This regulation can be seen only at a large spatial scale. Locally, N/P differences may be observed between seawater and biomass. This has maintained for a long time the debated question of the limiting nutrient in ocean ecosystems. The model proposed by Tyrell (1999) indicates that both N and P play a limiting role, but that the P control is sensitive at timescales longer than 100 000 years, while the N control is sensitive at climatic time scales (from some thousands to hundreds of thousands of years).

The present residence time of phosphate (PO_4^{3-}) in the ocean is about 80 000 years (Froelich et al. 1982). The input of dissolved phosphate to the ocean is mainly due to river discharge (Fig. 10.4), while the output is mainly due to sediment accumulation as organic phosphorus and apatite ($Ca_5(PO_4)_3OH$). Therefore, the P sources for the ocean are subjected to the erosion variability as well as to changes in the lithological nature of the eroded river basins. Although modulated by climate variability, such modifications are mostly sensitive to long-term geological changes such as orogenesis, sediment deposition in alluvial plains and deltas. The sedimentary flux (P sink) is subjected to modifications of the oceanographic settings also due to long-term phenomenon related to plate tectonics (basin opening or closing). The ocean phosphate reservoir is thus modified at a geological timescale because of positive or negative excursions of the phosphate source/sink global budget. At first glance, this is only due to the internal activity of the Earth. However, plate tectonics and the related phenomenon (volcanism, orogenesis, paleogeographic changes, etc.) are probably influenced

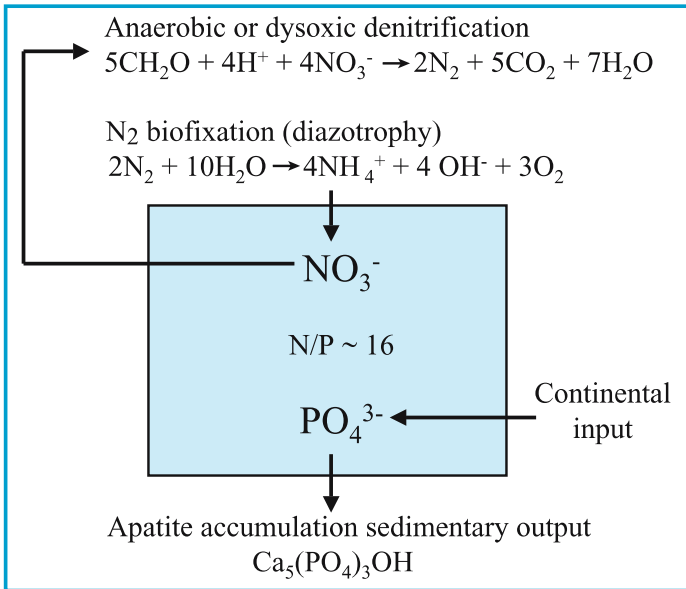


Fig. 10.4. The global ocean phosphate reservoir changes at a geological timescale under the influence of both the terrestrial input (river flux) and the sedimentary output. The global ocean nitrate reservoir changes at shorter timescales under the influence of both the N₂ biofixation flux (input) and the denitrification flux (output)

by marine biological productivity through sediment accumulation (isostasy) and subduction.

The present residence time of nitrogen as nitrate (NO₃⁻) in the ocean is only 8000 years (Schlesinger 1991). The main nitrate sources for oceans are continental organic matter re-mineralization and the N₂ biofixation (diazotrophy) by some groups of marine photosynthetic organisms (Fig. 10.4). In fact, the whole continental production of organic matter (except what is now produced anthropogenically using chemical fertilizers) is also based on diazotrophy, through soil bacteria, because this is the only source to renew the nitrogen in soil that can be assimilated by plants. The continental organic matter re-mineralization is thus a nitrate source to the ocean that results indirectly from diazotrophy. The nitrate sinks in the ocean are due to the biological uptake through marine production as well as to the bacterial transformation of nitrate to N₂ or N₂O (denitrification). The produced components are gases in surface conditions and may then be exchanged with the atmosphere. The bacterial denitrification takes place in anaerobic to dysaerobic conditions, occurring either in poorly ventilated layers of the water column in some oceanic areas (oxygen minimum zones or OMZ), or in sediments beneath the redox front. Even if continental vegetation or water discharge may change local nitrate sources, the global ocean nitrate inventory is mainly controlled by the balance between diazotro-

phy (source) and denitrification (sink). What are the factors that control this balance?

The most common photosynthesis uses nitrate as a nitrogen source. In the modern ocean, the dissolved nitrate content in seawater limits photosynthesis when $N/P \leq 16$, which implicitly means that only N and P are limiting amongst biogenic elements. This assumption is only valid at the global scale and requires further explanations. CO_2 is a non-limiting nutrient because the reservoir of dissolved inorganic carbon in the ocean, mainly as HCO_3^- , is very large. HCO_3^- may be easily transformed into CO_2 and CO_3^{2-} through displacement of the dissolved inorganic carbon equilibrium (more details are given in the following section). On the other hand, dissolved silicon, mainly as silicate $HSiO_3^-$, is limiting only for those primary producers that need to build siliceous tests, such as diatoms. Therefore, Si limits total primary production only when the role of the Si-requiring producers is significant or dominant, which depends on oceanographic conditions. For example, in high latitude areas, calcite bioprecipitation by calcifying producers is hampered by low temperatures so that diatom production may be locally favored and becomes dominant relative to the whole local production. Besides, in several other areas at lower latitudes, the producers that do not require Si (Coccolithophorida, Cyanobacteria, etc.) may easily replace silicifying producers to use available nutrients in the surface oceans, including the non-utilized nutrient coming from high latitude areas. Therefore, at the global scale, the production is essentially limited by N or P. Finally, a last limitation may come from the lack of micronutrients such as iron. Indeed, this element is generally insoluble in common redox conditions of the open ocean, which prevent it from being recycled through organic matter degradation in deep ocean and subsequent welling up to the ocean surface. As shown by Fe-fertilizing experiments, Fe strongly limits production in ocean areas that are weakly fed by terrigenous inputs. However, as previously discussed for Si, the Fe limitation impact is negligible at the global scale. Indeed, the nitrogen and phosphorous not utilized because of local Fe-limiting effects may be used later when the moving surface waters encounter more favorable conditions for terrestrial Fe input. When the budget of local effects is integrated at a global scale and at the timescale of ocean turnover (around 1000 years), the major nutrients (N and P) are probably the ultimate limiting nutrients for global Earth regulation. Let us now examine what their respective roles are in the context of the Tyrell's view (1999).

As previously mentioned, the ocean-dissolved P reservoir is mainly controlled by the continental inputs and the organic P and sedimentary apatite outputs (Fig. 10.4). It significantly changes only at a geological timescale ($> 1\text{Ma}$). Besides, the ocean dissolved N reservoir may change faster through the diazotrophy inputs and the denitrification outputs. Let us imagine a starting point with dissolved N/P equal to 16 (Fig. 10.5), which corresponds to the observed average situation of the present ocean (GEOSECS global data set, Tyrell 1999). If the dissolved phosphate reservoir increases, the dissolved N/P becomes lower than N/P of the dominant biological demand (16), which implies that N (as dissolved

nitrate) is the limiting nutrient. In such a case, the diazotrophic primary producers, because they are able to use N_2 instead of dissolved nitrate, are favored with respect to the nitrate users. Most of the produced organic matter is recycled so that, globally, diazotrophy transfers dissolved nitrogen from the N_2 reservoir to the NO_3^- reservoir. This transfer proceeds until the dissolved N/P reaches the value of the dominant biological demand (16). Conversely, if the dissolved phosphate reservoir decreases, nitrate is not limiting for the dominant biological demand so that the nitrate users are favored with respect to the diazotrophic species (the nitrate utilization energy yield is better than the N_2 biofixation one). While the nitrate users are favored and export organic matter down to the ocean interior, denitrification continues to proceed, which results partly in a nitrogen transfer to the atmosphere, through N_2 or N_2O degassing. Therefore, the nitrate ocean reservoir decreases down to a value that corresponds again to a dissolved N/P of 16.

This N/P regulation may have occurred since the biological and sedimentary nitrogen fluxes became dominant with respect to the abiotic nitrogen fluxes, which probably occurred very early after the onset of life on the Earth. However, both the present values of the biological demand and dissolved N/P is strongly dependent on the oxygen regulation that has occurred since about 500Ma. The N/P biological demand of Precambrian ecosystems, characterized by dominant

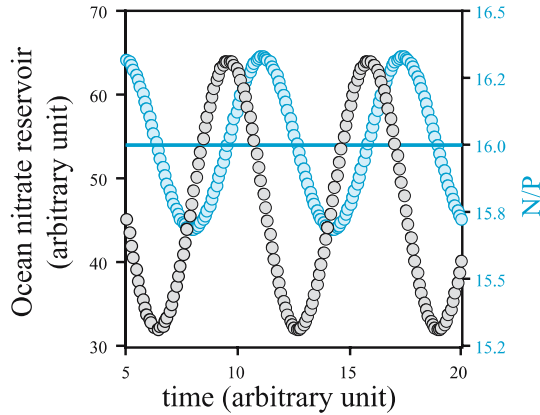


Fig. 10.5. The ocean-dissolved N/P is primarily affected by long-term changes in the ocean phosphate reservoir (see Fig. 10.4). As N/P deviates from the value of 16, corresponding to the biological demand of the marine phytoplankton (photosynthetic producers), the ecosystem response favors either the dominant nitrate users ($N/P > 16$) or the primary producers that are able to assimilate N_2 ($N/P < 16$). In the latter case, the nitrogen fixation flux into the ocean (or diazotrophic flux) becomes more important than the nitrate loss due to denitrification, which tends to increase the N/P ratio back to 16. Conversely, when the nitrate users are favored, the denitrification output exceeds the diazotrophic input, which, similarly, tends to lower the N/P back to 16. Although not always limiting, nitrogen is the main regulating nutrient in the ocean

anoxia (before 2.1 Ga) or by low availability of O_2 (between 2100 and 540 Ma, Fig. 10.2) should have been very different because the role of ammonium (NH_4^+) assimilation in the nitrogen cycle was much more important. Therefore, the N/P regulation at short timescales has probably occurred for a long time but with different values of the N/P ratio controlled by the steps of the oxygen regulation as well as by geological contingencies affecting the ocean P reservoir. In this regulation, P is the *ultimate* limiting nutrient. This means that the ocean nitrogen reservoir adjusts to it to maintain the dissolved N/P equal to that of the biological demand. It is generally believed that the long-term variability of the oceanic P and N reservoirs depends only on the phosphorous ratio between the source (river discharge) and the sink (sedimentation). However, as previously mentioned, a biological retroaction may not be excluded through the role of the biogenic sedimentation on the plate tectonic activity.

10.3 What About the Atmospheric CO_2 ?

The pre-industrial CO_2 content in the atmosphere (280 ppm), or even the present one (370 ppm) is nearly nothing in comparison with the CO_2 content of the primitive atmosphere where CO_2 was probably the dominant gas ($> 950\,000$ ppm). Is the atmospheric CO_2 content decrease so ineluctable that the carbon supply to continental photosynthesis could be endangered in the future? Of course, addressing this question is surprising in the human development context where the excess of atmospheric CO_2 more than its deficit is a problem. The question is nevertheless important for the global ecosystem knowledge, which is necessary to improve the reliability of the predicting climate models.

We will consider that the natural variability of the global ecosystem corresponds to the pre-industrial situation because most of the anthropogenic effect on atmospheric CO_2 has occurred for less than one century. However, we should keep in mind that the true natural variability finished at the onset of the Neolithic period (10 000 years ago), when the first agricultural practices began. The non-natural variability observed for 100 years led to a present atmospheric CO_2 concentration, which is much higher than any other one observed for the last 5 Ma, but, nevertheless, much lower than during older periods such as the Cretaceous.

Few measured data are available to quantitatively document the atmospheric CO_2 variability through geological times, avoiding speculative working hypotheses. The most reliable data are for sure the Vostok or Dome C Antarctic ice core records, which cover only the last 420 000 years (Fig. 10.6, Petit et al. 1999) or the last 740 000 years (EPICA 2004), respectively. Although limited to the late Quaternary, these data provide the fundamental information that the atmospheric CO_2 content of the modern global ecosystem seems to be naturally regulated at a timescale longer than the glacial-interglacial fluctuation. In particular, they show that the pre-industrial atmospheric CO_2 content is not the

lowest one experienced during the modern global ecosystem history and, therefore, that natural regulating mechanisms proceed. Such phenomenon acts at short timescales (climatic variability forcing, mostly orbital), but in a context where the long-term average atmospheric CO₂ has been relatively stabilized.

What are the CO₂ regulating phenomena? The present knowledge identifies some of them, but cannot yet measure their relative importance (Pedersen and Bertrand 2000; Archer and Winguth 2000; Sigman and Boyle 2000). Although the occurrence of glacial-interglacial fluctuations has been a strong feature of the Earth's climate for more than 2Ma, as well as during some older periods, it is far to be a permanent feature on the planet. It depends, among other factors, on the geographical distribution of the continents and on the Earth's rotational axis. For example, the Antarctic ice cap would not exist if an insular continent was not presently centered on the South Pole. Moreover, this ice cap would not be stable if the continent was not be isolated by the peri-Antarctic ocean circulation and the related atmosphere dynamic. Amongst the proposed CO₂-regulating mechanisms, some are strongly linked to the modern Earth's climate situation, such as air-sea exchanges or modifications of the global ocean circulation, while some others are more independent of glacial-interglacial fluctuations and may have occurred for much longer periods such as silicate weathering and sedimentation of organic matter and carbonates.

Hereafter is an example of a mechanism that is typically linked to the glacial-interglacial fluctuation, proposed by Broecker and Henderson (1998). The Vostok Antarctic ice core record shows that the atmospheric dust flux, coming from Patagonia, cease abruptly at the onset of deglaciations. Because the dust contains iron, which is a limiting factor in the Southern Ocean, they stimulate the diazotrophic production in this ocean. When the dust flux ceases, the regional diazotrophic production and consequently the related global ocean nitrate inventory, decreases. The global ocean production also progressively decreases, weakening the ocean CO₂ pump with respect to the atmosphere. During such

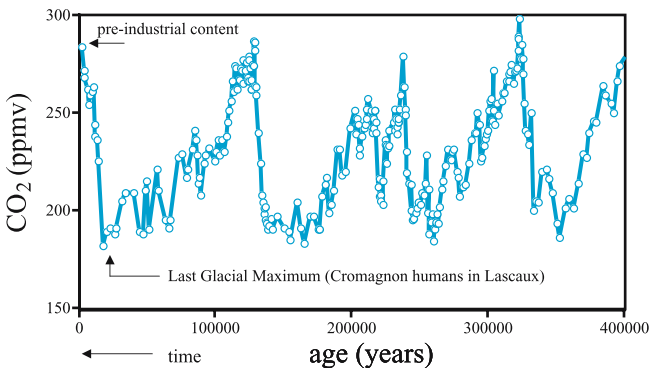


Fig. 10.6. Evolution of the atmospheric CO₂ partial pressure over the last 400 000 years, as recorded by the Vostok ice core, Antarctica (Petit et al. 1999)

periods, the CO₂ sources dominate against CO₂ sinks so that the atmospheric CO₂ content increases as observed. This hypothesis has been modulated by Bopp et al. (2003) which results show that dust flux variability to Austral Ocean cannot explain more than 30% of the atmospheric CO₂ variability. Moreover, an uncertainty is that beyond the observed facts and data, the proposed regulating mechanisms use the concept of source and sink generally without specifying the considered timescale. Indeed, such a concept needs to be better defined when interpreting the observations. First, it is very important to clearly distinguish what is the immediate consequence of a phenomenon and what is its delayed consequence. For the carbon cycle, the immediate effect of a decrease in productivity is to reduce the transfer of particulate carbon, either as organic matter or biocarbonate, from the surface down to the ocean interior (below the surface mixing layer). This transfer may be more or less efficient depending on the considered oceanic regions and the eustatic conditions (Bertrand et al. 2000). For example, eutrophic conditions favor the organic carbon export efficiency through large faecal pellets or aggregates. This effect is enhanced during low stand sea level, when the particles are more or less recycled over the continental shelf, and rapidly exported from the surface ocean. Because the dissolved CO₂ partial pressure in the surface ocean is lowered or increased, through the dissolved inorganic carbon equilibrium, when organic carbon and carbonate are exported respectively, the transfer down to the ocean interior induces immediate consequences for the carbon exchanges between ocean and atmosphere. However, the whole budget must also take into account the fate of particles exported down to the ocean interior. Both the organic matter re-mineralization and carbonate dissolution through the inorganic carbon equilibrium change the chemical characteristics of deep water masses. Through upwelling or diffusion, the new characteristics may be transferred back to the surface, which cancels or modulates the previous effects of particle exportation. The understanding of ocean-atmosphere CO₂ exchange is difficult because all the potential forcing factors must be taken into account, as well as their inter- or retroactions, over a period of time not shorter than the ocean turnover (1000 to 2000 years). Meeting these requirements is only possible by using a computed global modeling approach. The role of the field and experimental observation focuses the model on the most important phenomenon, and to realistic core parameters, to propose realistic forcing input scenarios, and to control modeled outputs. We are still far from such a situation, although the scientific community is rapidly progressing towards the concept of a global model including the long-term biogeochemical cycles (Berner 2004). Important compartments such as the whole carbon cycle, including the role of continental margins, are still extremely rough or even absent. The theoretical global approach is therefore important to refine models and to elaborate future program strategies and regional implementation plans.

The ocean surface CO₂ partial pressure (p_{CO_2}) and the atmospheric p_{CO_2} tend to balance themselves, so that any regional or global change in the ocean, dissolved inorganic carbon equilibrium (Fig. 10.7) plays an important role in

controlling atmospheric p_{CO_2} . However, this balance may be delayed. Indeed, if the p_{CO_2} perturbation takes place in the ocean interior, for example bottom carbonate dissolution due to a re-organized deep ocean circulation, the related new chemical properties of seawater must be transferred up to the surface before they can induce an ocean-atmosphere carbon exchange. This transfer may occur through either thermohaline circulation and subsequent upwelling, or progressive upward diffusion. The main phenomenon that are active with respect to the dissolved inorganic carbon equilibrium are the photosynthesis (CO_2 consumption), the respiration (CO_2 production), the CaCO_3 bioprecipitation (dissolved CO_3^{2-} consumption) and the CaCO_3 dissolution (dissolved CO_3^{2-} production). Each may interact with the others. For example, the deep or bottom heterotrophic respiration depends on sinking organic matter that was previously photosynthe-

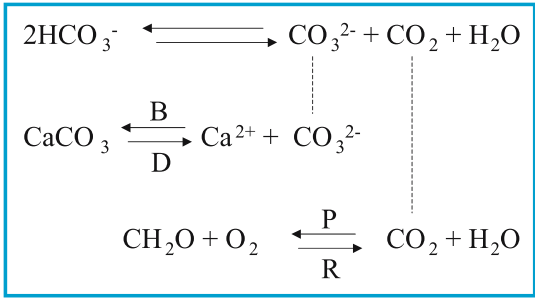


Fig. 10.7. The photosynthesis (P), the respiratory organic matter re-mineralization (R), the carbonate bioprecipitation (B), and the carbonate dissolution (D) contribute to change the dissolved inorganic carbon equilibrium in the ocean ($\text{HCO}_3^-/\text{CO}_3^{2-}/\text{CO}_2$). At long timescales, the cumulated budget of the different local fossilization of sedimentary carbonates ($B > D$). The response of the ocean system compensates these two effects by converting HCO_3^- into CO_2 and CO_3^{2-} . However, because the two-fossilization fluxes are never synchronously equivalent, this response induces either a CO_3^{2-} oversaturation (when organic matter fossilization is dominant), which favors retroactively carbonate bioprecipitation, or a CO_2 oversaturation, which implies an increase of atmospheric CO_2 partial pressure (when carbonate fossilization is dominant). Because the HCO_3^- reservoir is much larger than the dissolved CO_3^{2-} and CO_2 reservoirs, each of these two processes is limiting for the other one within a given timescale (corresponding to the time necessary to reduce significantly the HCO_3^- reservoir). Of course, the geological recycling of carbon and calcium through volcanism and continental weathering limits ultimately this ocean chemistry regulation. Because the geologically recycled carbon comes from subducted or weathered sediments, it is probable that the carbon sedimentation to carbon recycling flux ratio is itself regulated at very long timescales. Let us emphasize that the volcanic activity, on which the long-term carbon recycling is partly based, may occur as catastrophic abrupt events, impacting heavily the global ecosystem and stimulating the short-term regulating processes

sized in the light-exposed surface water (photic layer). In turn, this respiration releases CO_2 that modifies the dissolved inorganic carbon equilibrium and favors the dissolution of previously precipitated CaCO_3 . These interactions explain why the depth at which the sedimentary carbonate are entirely dissolved (carbonate compensation depth or CCD) differs from one ocean area to another. The CCD is deeper in the Pacific Ocean ($\approx 4000\text{m}$) because its interior is less ventilated than the Atlantic Ocean interior where CCD is around 3500m . Indeed, deep heterotrophic respiration requires not only organic matter, but also oxygen. Consequently, the sedimentary carbonates are overall, better preserved in the Pacific than in the Atlantic.¹

Measurements have shown that some ocean areas act as local CO_2 sources although they are highly productive (Pedersen and Bertrand 2000). This is the case for the East Pacific equatorial upwelling area, induced by the equatorial low-pressure system (or the inter-tropical convergence zone, ITCZ) to which the trade wind dynamics are related. How is this possible? Two complementary explanations account for this phenomenon. First, the p_{CO_2} of the up-welled water may be high. Second, the CaCO_3 bioprecipitation (tests, skeletons) and subsequent transfer down to the ocean interior result in CO_2 production in the surface water through the displacement of the dissolved inorganic carbon equilibrium as a response of dissolved CO_3^{2-} consumption (Fig. 10.7). If the CO_2 input of these two mechanisms is greater than the CO_2 output due to the export flux of photosynthesized organic matter down to the ocean interior, then the surface water p_{CO_2} becomes higher than the atmospheric p_{CO_2} and the area acts as an immediate CO_2 source for the atmosphere. However, if the fate of particles that have been exported down to the ocean interior is considered over the time of the ocean turnover, the assessment of the impact of the east equatorial Pacific area may be significantly different. If sinking CaCO_3 particles are deposited below the local CCD, they will be entirely dissolved, inducing a decrease in the deep-water CO_2 content (Fig. 10.7). Through subsequent thermohaline circulation or upward diffusion, this low CO_2 content property will be transferred up to the surface and act as a CO_2 pump for atmosphere, which may cancel or moderate the initial CO_2 source. Reversely, if CaCO_3 particles are deposited over the local CCD and, hence, preserved as sedimentary carbonate, then the initial effect is also preserved.

Up to now, most of the hypotheses proposed to account for atmospheric p_{CO_2} variability during the late Quaternary are exclusively based on transfer mechanisms between the atmosphere, the surface ocean, and the deep ocean. The preserved sedimentary fluxes are considered negligible with respect to the

¹ We should keep in mind that the CCD is meaningless for bottom carbonate dissolution in the continental margins. Indeed, these areas are often highly productive because of intense nutrient fertilization (eutrophic areas) and induce important organic matter fluxes down to the seafloor. The subsequent bottom organic matter degradation may release locally enough CO_2 to induce carbonate dissolution at depth above the CCD of the corresponding open ocean

other fluxes at such a timescale. But, as strongly suggested by regional studies (Bertrand 2002), it is probable that the global ratio between the preserved sedimentary organic carbon and the preserved carbonate may, to some degree, control the atmospheric p_{CO_2} variability, not only at long (geological) timescales, but also at climatic timescales.

10.4 Towards a Global Biogeochemical Regulation (Homeostasy)

The life on the Earth is mainly based on the availability of water and carbon. Its diversity uses the countless possibilities of the organic chemistry in aqueous solution. Would its longevity be related to the inorganic carbon chemistry in aqueous solution?

By taking into account the knowledge and hypotheses available from the literature, this section attempts to give a synthetic and hypothetical view on how the atmospheric CO_2 regulation settles in relation with the oxygen and nutrient settlements. Of course, such a speculative and global view requires simplifying the reality. In particular, it does not take into account the up to now non-measurable relative influences of the different areas (eutrophic/oligotrophic, low/high latitudes, above CCD/below CCD, etc.). It is based on the physicochemistry of the inorganic carbon in aqueous solution, as well as on the reasonable postulate of the occurrence of different processes since early life on the Earth (Pinti 2003): a photosynthetic production in surface ocean; a plate tectonic activity and a related geological recycling of sediments.

First, we emphasize that if the proportions of the different species of dissolved inorganic carbon in aqueous solution depend on the pH (Bjerrum diagram, Fig. 10.8), the reverse is true. Indeed, CO_2 may turn partly into H_2CO_3 , HCO_3^- and CO_3^{2-} , or reversely. The dissociation of the carbonic acid (H_2CO_3) and of the bicarbonate anion (HCO_3^-) lowers the pH (Zeebe and Wolf-Gladrow 2001). Adding or removing CO_2 results therefore in a pH decrease or increase, respectively. In the physicochemical conditions of the present ocean, the dissolved CO_2 content is two orders of magnitude lower than the HCO_3^- content, which is a weakly dissociated acid. This means that the added or removed CO_2 amount must be considerable in order to have a significant relative impact on the HCO_3^- content and consequently to change the pH. This corresponds to the so-called buffering capacity of the ocean carbonate system. Moreover, an excessive pH decrease or increase, due to the exhaustion of the ocean own buffering capacity, may be buffered through exchanges with the sedimentary reservoir. The sedimentary preservation of bioprecipitated biocarbonates tends to compensate for an excessive increase of pH resulting in a higher CO_3^{2-} content (Fig. 10.8), which also compensates for a CO_2 deficit. Reversely, the dissolution of biocarbonates tends to compensate for an excessive decrease of pH and the related increase of CO_2 content. Although the dissolved inorganic carbon system may

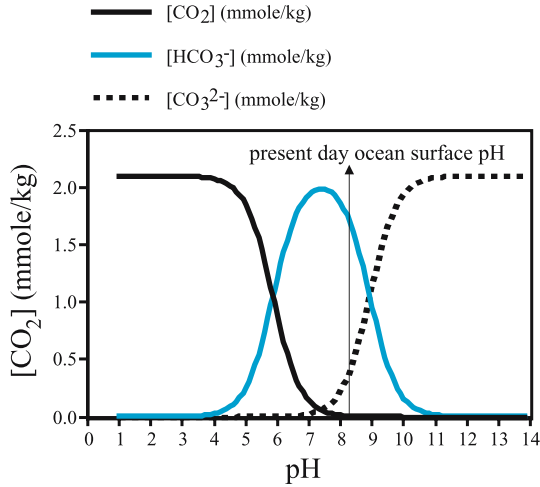


Fig. 10.8. The Bjerrum diagram (named after its author) predicts the distribution of the main dissolved inorganic carbon species (HCO_3^- , CO_3^{2-} and CO_2) in aqueous solution as a function of pH. Here, the distribution curves are calculated for a common situation of the ocean surface where the total inorganic carbon content (TIC) is 2.1 mmole/kg, the atmospheric pressure is 1 bar, the temperature is 25 °C and the salinity is 35‰. At the average pH of the present surface ocean (about 8.2), the dissolved inorganic carbon reservoir occurs mainly as bicarbonate (HCO_3^-). At timescales shorter than that required for geological sediment recycling, any ocean regulation is based on this buffering capacity

be locally and temporarily perturbed, it is globally and on the long-term extremely stable, at least while HCO_3^- is very dominant with respect to CO_2 and CO_3^{2-} , which is the case as long as the sedimentary CaCO_3 reservoir is not exhausted.

Another aspect important to consider is the relative proportions of the carbon reservoirs. The modern Earth sedimentary carbon reservoir is 2300 times the ocean-dissolved inorganic carbon reservoir (mainly as HCO_3^-), itself being 65 times the atmospheric CO_2 carbon reservoir (Chameides and Perdue 1997). The terrestrial biosphere contains a quantity of carbon of the same order of magnitude than the atmospheric CO_2 . Any decrease of atmospheric p_{CO_2} not originating from the ocean is compensated by a rapid balancing with the dissolved p_{CO_2} of the surface ocean, which consumption may be easily buffered by the huge HCO_3^- reservoir through the dissolved inorganic carbon equilibrium. Therefore, even ignoring the volcanic and anthropogenic CO_2 inputs, the atmospheric p_{CO_2} cannot rapidly decrease to such extent that the continental photosynthesis would be endangered, at least within the present pH conditions of the ocean. At longer timescales, the ocean-dissolved inorganic carbon chemistry is even buffered by removing part of the huge sedimentary carbon reservoir through CaCO_3 dissolution. As we can see, the main regulating processes of the

Earth system occur within the ocean. That is why the following discussion will focus on them.

The average pH of the ocean surface is about 8.2. In such conditions, the dissolved inorganic carbon chemistry predicts that the dominant carbon chemical form is bicarbonate (HCO_3^-) while CO_2 and CO_3^{2-} are only minor carbon occurrences (Fig. 10.8). In addition, the Bjerrum diagram shows that the situation is exceptional because it corresponds to a narrow pH range, between a large range where CO_2 is dominant ($\text{pH} < 6$), and a large range where CO_3^{2-} is dominant ($\text{pH} > 9$). As previously mentioned, such a situation corresponds to a stability sink as long as the HCO_3^- is not significantly modified (buffering effect). With respect to the Earth's regulation, it is important to know whether this situation constitutes just a step or an outcome of the biogeochemical evolution of the ocean. This question was previously addressed by Sillèn some decades ago (Sillèn 1961, 1967). This author concluded that the physicochemistry of the present ocean, and hence, its pH, is mainly controlled by regulating exchanges between the solid and dissolved reservoirs of silicates, and that this equilibrium was probably reached very early in the Earth's history. The supplementary knowledge that is available today, especially the probable hypothesis of a CO_2 -dominated early atmosphere before the onset of the early ocean, leads to re-assess such a conclusion. Especially, the question arises to know whether the oceanic carbon/sedimentary carbon system may have played a regulating role as, or possibly more important than the system of silicates in the Earth's history. In other words, is the global ocean buffering capacity, which is a climatic timescale regulating mechanism of the Earth system, a stable feature at a geological timescale?

When any instrumental or numerical experiment is unaccessible, it is only possible to propose some working hypothesis based on theoretical mind experiments. First, let us establish what we consider as the primary force in these experiments. The ocean receives directly or indirectly (through the continental weathering) carbon from the volcanic CO_2 degassing, thus in other words from the plate tectonic activity. This is a primary force. We must consider any other input (sedimentary carbonate dissolution, sedimentary organic matter remineralization, etc.) as just a consequence of a change in ocean conditions, therefore as a retroactive response to the primary force. In the argument below, we set apart the possible regulation role of silicates described by Sillèn (1967).

For the first experiment, we suppose that the ocean is subjected to an increase of CO_2 input due to an enhanced volcanic activity. The CO_2 in excess is largely transformed into HCO_3^- with, consequently, a decrease in the dissolved CO_3^{2-} reservoir that is partly compensated by dissolving sedimentary CaCO_3 (Fig. 10.7). If the volcanic primary force lasts long enough to exhaust the buffering capacity of the sedimentary CaCO_3 , then the ocean conditions would drastically change, especially pH would strongly decrease, ultimately reaching that of the primitive ocean in equilibrium with an atmosphere dominated by CO_2 . However, this very long-term issue is unrealistic because the volcanic CO_2 input itself originated from the sedimentary reservoir through geological recycling

(plate tectonic-related subduction). Thus, this input would come back to lower values much before the complete consumption of the huge sedimentary CaCO_3 reservoir. Moreover, at a geological timescale, volcanic CO_2 in excess would be also partly used for the weathering of continental aluminosilicates (Berner et al. 1983; Berner 2004), especially of the basalts that would be unavoidably released by the enhanced volcanic activity. This alteration transforms atmospheric CO_2 into dissolved HCO_3^- , which is then transferred into the ocean by rivers. In the ocean, this HCO_3^- input is buffered through the dissolved inorganic carbon equilibrium, converting HCO_3^- into CO_2 and CO_3^{2-} (Fig. 10.7). In the end, in the pH conditions of the modern ocean, CO_3^{2-} in excess is converted to sedimentary CaCO_3 that tends to balance the previously described effect of a direct CO_2 transfer to the ocean.

For the second experiment, we have just to suppose an opposite primary force, that is to say a decrease of CO_2 input due to a weakened volcanic activity. All things being equal, the dissolved CO_2 deficit would be buffered by converting HCO_3^- into CO_2 and CO_3^{2-} , with the consequence of an increase of the sedimentary CaCO_3 reservoir to compensate the excess in CO_3^{2-} . In the very long-term, the increase of the sedimentary CaCO_3 reservoir would induce an enhanced volcanic CO_2 release because of a greater flux of subducted sedimentary carbon. This constitutes the retroactive limit of the initial primary force.

These two mind experiments show that the ocean physicochemical conditions, especially the pH, cannot durably deviate from the present conditions without a deep modification of the sedimentary carbon reservoir. Moreover, at the timescale of the geological sediment recycling, such a modification induces a negative retroaction by limiting the initial forcing. This long-term regulation of the carbon cycle was recently extensively described by Berner (2004). It is mainly modulated by the tectonic activity and the related processes (volcanism, metamorphism, orogenesis, erosion, sedimentation, etc.). As an example, the rapid uplift of a chain of mountains, such as the present Himalaya, induces steep slopes that enhance the weathering rate of silicates and the related uptake of atmospheric CO_2 (Gaillardet et al. 1999). But the reduction of the greenhouse effect due to this uptake leads to a cooling effect that, retroactively, lessens the weathering rate of silicates. Moreover, this weathering rate may have been modulated by processes such as the evolution of the continental vegetation throughout the Phanerozoic. Indeed, the trees, because of their vast rootlet systems, accelerate the chemical erosion more than lichens and mosses (Moulton et al. 2000).

Beyond this regulating power of the coupled volcanic activity and sedimentation system, it remains to determine whether the modern Earth is close to the attractive regulated carbon Earth situation (or the climate “attractor”, in the Lorentz’s sense). In the following discussion, we will describe the evolution of the global carbon Earth’s towards this attractor by the evolution of the relative proportions of the sedimentary and ocean carbon reservoirs, of the redox com-

position of the sedimentary carbon reservoir (organic matter/carbonate), and of the pH-related composition of the ocean-dissolved inorganic carbon reservoir ($\text{CO}_2/\text{HCO}_3^-/\text{CO}_3^{2-}$).

From the available literature, it is found that most of the primitive CO_2 Earth's degassing would have occurred before the accumulation of the sedimentary reservoir, so that the primitive atmosphere would have been mostly constituted of CO_2 , and largely overpressured when compared with the present atmosphere. Indeed, the primitive atmospheric pressure is estimated between 40 and 210 bars (Pinti 2003), which corresponds to 40 to 210 times the present atmospheric pressure (1 bar). Assuming this pressure estimation, the primitive ocean would have probably been much more acid than today, hindering the precipitation of CaCO_3 as well as its sedimentary accumulation. However, this is not obvious. Indeed, the ocean-dissolved p_{CO_2} must have been very high to balance at least 40 bars of atmospheric CO_2 . Considering the Bjerrum diagram at a much higher atmospheric pressure than today (Zeebe and Wolf-Gladrow 2001), this is compatible with an HCO_3^- and CO_3^{2-} ocean on the same order of magnitude as today, at pH as low as 5, or even less. Moreover, the evolution of the solid/dissolved silicate system might have contributed to limit the acidity of the ocean (Sillén 1967). Therefore, the carbonate precipitation and sedimentary accumulation within the primitive ocean conditions cannot be excluded. Also, Ca^{2+} and Fe^{2+} and some other cations that are soluble in anoxic conditions, should have been available in the primitive sea water. Besides, the carbon could accumulate easily as organic matter because of the primitive anoxia. Therefore, this early sedimentary accumulation of both organic matter and, possibly, carbonates may have reduced the primitive ocean and atmospheric carbon reservoirs as long as it was not fully compensated by the CO_2 input due to the geological sediment recycling. Based on the present rate of plate tectonic activity and sedimentation, reaching this compensation would require more than 100Ma. As discussed earlier in this chapter, the increase of the sedimentary organic carbon reservoir means that O_2 resulting from photosynthesis is not entirely recycled by either oxic or anoxic respiration and, hence, is accumulating in the environment or is consumed by oxidation of other reduced chemical species, mainly sulfides. The O_2 accumulation in the Earth's environment has led the evolving life towards the development of aerobes, which have probably contributed to slow, then to stop, the net flux of sedimentary organic carbon (accumulated re-mineralized geologically recycled). In such new conditions, the $C_{\text{organique}}/C_{\text{carbonate}}$ ratio of the sediment reservoir must have decreased progressively, down to a value where this evolution was limited by the respiration of deep aerobes. Indeed, this respiration induces an elevated CO_2 content of deep water that increases its corrosivity with respect to sedimentary carbonates (Fig. 10.7). Finally, we can consider that the situation of both the modern ocean and atmosphere results from the evolution of the sedimentary reservoir, which sources and sinks are ultimately controlled by the physicochemical conditions of the ocean and the volcanic CO_2 input (geological recycling), respectively. Temporary perturbations may occur (climatic

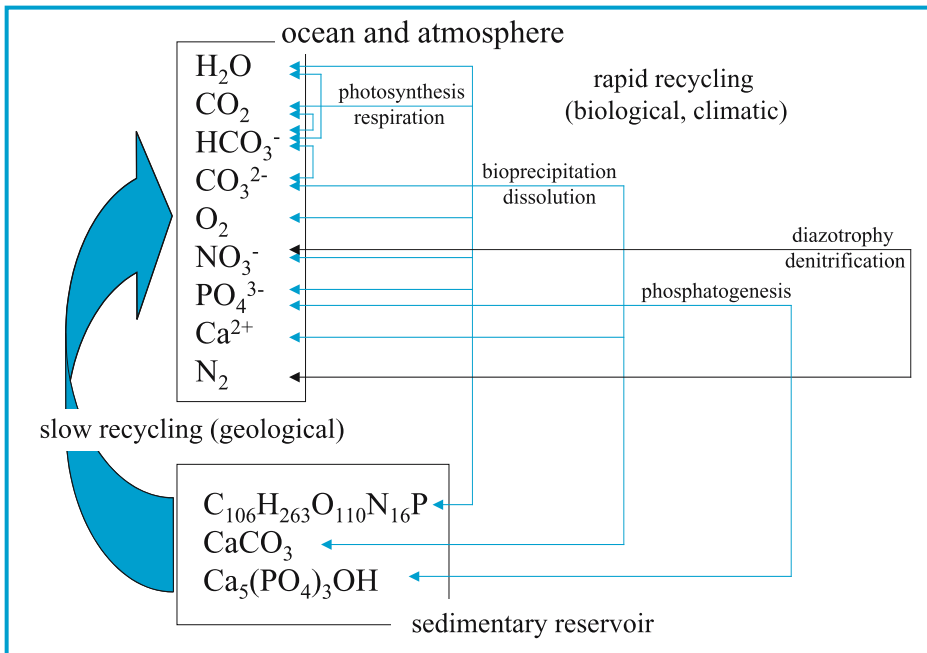


Fig. 10.9. The Earth's life induces important fluxes between the ocean and sediment reservoirs. Acting together with the longer term sediment recycling, powered by the plate tectonic activity (volcanism, mountain build-up and subsequent continental weathering), these processes constitute a powerful cybernetic machinery. This machinery is able to endure a large range of perturbations either external (asteroids, insolation) or internal (climatic response to the geographical distribution of continents and oceans, threshold effects, etc.). Most of the Earth's history has consisted of approaching equilibrium between the sediment and ocean reservoirs, mainly based on the stability conditions of the dissolved inorganic carbon in aqueous solution. Although this equilibrium may not be fully reached, some robust regulations observed in the modern Earth's ecosystem have probably occurred for hundreds of millions of years (O_2 , N/P , CO_2)

or volcanic abrupt events or episodes, etc.), such as, for instance, the conditions that would have allowed the onset and breaking-up of global glaciations, the so-called snowball Earth episodes (Goddéris et al. 2003; Donnadieu et al. 2004; Ramstein et al. 2004). However, at timescales longer than these possible perturbations, the main global features of the modern Earth system (atmospheric pressure, O_2 abundance, ocean pH) seem to result from a robust regulation controlled in large extent by the related evolution of both the sedimentary carbon reservoir and the proportions of the respiratory metabolisms (Fig. 10.9).

Within this context, the anthropogenic perturbation should be assessed differently as the short or the long-term effects are considered. The CO_2 input

due to fossil fuel combustion may be compared to the input of an important volcanic episode. It has no short-term effect on the sedimentary reservoir and will be easily counterbalanced by the global regulating ocean system even if the transient stage will probably induce considerable perturbations for the human society. The consequences of the biodiversity reduction, especially on that of multicellular eukaryotes, will be more durable. They should not affect too much the long-term regulatory mechanisms, which are essentially biological, and based on the huge metabolic activity of the dominant biomass, constituted of both the prokaryotic aerobes and anaerobes. Here again, the humans endanger only the “Garden of Eden” but not the Earth’s ecosystem, as mentioned by Emmanuel Kant in his *Nature General History*. Let us hope the humans will be aware soon enough!

Acknowledgements

I wish to thank the organizers of the Exobio 2003 CNRS Thematic School, with special mention to Muriel Gargaud, for inviting me to give a talk. I also thank Gilles Ramstein, Robert A. Berner, Philippe Claeys and an anonymous colleague, as reviewers, for helping me to improve the manuscript. The preparation of both this conference and this chapter has been an opportunity to draw an evolutive view linking what we assume for the primitive or ancient global Earth’s ecosystem to what we know of its modern regulation. I have no doubt that this view is very broad and needs to be pushed further by interested colleagues. I think this subject constitutes a fertile research field for the future of exobiology.

References

- Archer, D., and Winguth, A., (2000). What caused the glacial/interglacial atmospheric $p\text{CO}_2$ cycles, *Rev. Geophys.*, **38**(2), 159–189.
- Bertrand, P. (2002). Peut-on évaluer le rôle de la sédimentation dans l’effet de serre à l’échelle de temps des cycles orbitaux? *C. R. Geosci.*, Série IIa, **334**, 521–528.
- Bertrand, P. and Lallier-Vergès, E. (1993). Past sedimentary organic matter accumulation and degradation controlled by productivity, *Nature*, **364**, 786–788.
- Bertrand, P., Pedersen, T.F., Martinez, P., Calvert, S.E., and Shimmiel, G. (2000). Sea level impact on nutrient cycling in coastal upwelling areas during deglaciation: evidence from nitrogen isotopes, *Global Biogeochem. Cycles*, **14**(1), 341–355.
- Berner, R.A. (2004). *The Phanerozoic Carbon Cycle*. Oxford University Press, New York, Oxford.
- Berner, R.A., and Canfield, D.E. (1989). A new model for atmospheric oxygen over phanerozoic time, *Am. J. Science*, **289**, 333–361.
- Berner, R.A., Lasaga, A.C., and Garrels, R.M. (1983). The carbonate-silicate geochemical cycle and its effect on atmospheric carbon dioxide over the past 100 million years, *Am. J. Sci.*, **283**, 641–683.
- Bopp, L., Kohfeld, K., Le Quéré, C. and Aumont, O. (2003). Dust impact on marine biota and atmospheric CO_2 during glacial periods, *Paleoceanography*, **18**, 2, 1046.

- Broecker, W.S., and Henderson, G.M. (1998). The sequence of events surrounding Termination II and their implications for the cause of glacial-interglacial CO₂ changes, *Paleoceanography*, **13**(4), 352–364.
- Chameides, W.L. and Perdue, E.M. (1997). *Biogeochemical Cycles*, Oxford University Press, New York, Oxford.
- Cope, M.J. and Chaloner, W.G. (1985). Wildfire: an interaction of biological and physical processes. In: Tiffney, B.H. (ed.) *Geological Factors and the Evolution of Plants*, pp. 257–277. Yale University Press, New Haven.
- Donnadiou, Y., Godd ris, Y., Ramstein, G., N d lec, A. and Meert, J. (2004). A snowball Earth climate triggered by continental break-up through changes in runoff, *Nature*, **428**, 303–306.
- EPICA, Community Members (2004). Eight glacial cycles from an Antarctic ice core, *Nature*, **429**, 623–628.
- Froelich, P.N., Bender, M.L., Luedtke, N.A., Heath, G.R. and De Vries, T. (1982). The marine phosphorus cycle, *Amer. J. Sci.*, **282**, 474–511.
- Gaillardet, J., Dupr , B., Louvat, P., Allegre, C. J. (1999). Global silicate weathering and CO₂ consumption rates deduced from the chemistry of large rivers, *Chem. Geol.*, **159**, 3–30.
- Godd ris, Y., Donnadiou, Y., N d lec, A., Dupr , B., Dessert, C., Grard, A., Ramstein, G. and Fran ois, L.M. (2003). The Sturtian ‘snowball’ glaciation: fire and ice, *Earth Planet. Sci. Lett.*, **211**, 1–12.
- Kah, L.C., Lyons, T.W., Franck, T.D. (2004). Low marine sulphate and protracted oxygenation of the Proterozoic biosphere, *Nature*, **431** 834–838
- Kupiec, J.-J. and Sonigo, P. (2000). Ni Dieu ni g ne: pour une autre th orie de l’h r dit . Editions Seuil, Paris.
- Lenton, T.M., and Watson, A.J. (2000a). Redfield revisited. 1. Regulation of nitrate, phosphate, and oxygen in the ocean, *Global Biogeochem. Cycles*, **14**(1), 225–248.
- Lenton, T.M., and Watson, A.J. (2000b). Redfield revisited. 2. What regulates the oxygen content of the atmosphere?, *Global Biogeochem. Cycles*, **14**(1), 249–268.
- Lovelock, J. (1990). Les  ges de Ga ia. Robert Laffont, Paris.
- Margulis, L. and Sagan, D. (1989). L’Univers bact riel. Editions Albin Michel, Paris.
- Massabuau, J.-C. (2003). Primitive, and protective, our cellular oxygenation status?, *Mech. Ageing Develop.*, **124**, 857–863.
- Moulton, K.L., West, J. and Berner, R.A. (2000). Solute flux and mineral mass balance approaches to the quantification of plant effects on silicate weathering. *Amer. J. Sci.*, **300**, 539–570.
- Pedersen, T.F., and Bertrand, P. (2000). Influences of oceanic rheostats and amplifiers on atmospheric CO₂ content during the Late Quaternary, *Quaternary Sci. Rev.*, **19**, 273–283.
- Petit, J.R., Jouzel, J., Raynaud, D., Barkov, N.I., Barnola, J.-M., Basile, I., Benders, M., Chappelaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotlkyakov, V.M., Legrand, M., Lipenkov, V.Y., Lorius, C., P pin, L., Ritz, C., Saltzman, E., and Stievenard, M. (1999). Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica, *Nature*, **399**, 429–436.
- Pinti, D. (2003). La formation des oc ans. In: Gargaud, M. (ed.) *Les traces du vivant*, pp. 39–57. Presses Universitaires de Bordeaux, France.
- Ramstein, G., Donnadiou, Y. and Godd ris, Y. (2004). Les glaciations du Prot rozo ique, *C. R. Geosci.*, **336**, 639–646.

- Redfield, A.C. (1958). The biological control of chemical factors in the environment, *Amer. Scien.*, **46**, 205–221.
- Schlesinger, W.H. (1991). Biogeochemistry: an analysis of global change. Academic, San Diego.
- Sigman, D.M., and Boyle, E.A. (2000). Glacial/interglacial variations in atmospheric carbon dioxide, *Nature*, **407**, 859–869.
- Sillén, L.G. (1961). The physical chemistry of sea water. In: Sears, M. (ed.) Oceanography, *Assoc. Adv. Sci.*, **67**, 549–581.
- Sillén, L.G. (1967). The ocean as a chemical system, *Science*, **156**, 1189–1197.
- Stengers, I. (2003). Gaïa la chatouilleuse, *La Recherche*, Hors Série **11**, 36–39.
- Stow, D., Huc, A.Y., and Bertrand, P. (2001). Depositional processes of black shales in deep water. *Marine Petroleum Geology*, **18**, 491–498.
- Tyrell, T. (1999). The relative influences of nitrogen and phosphorus on oceanic primary production, *Nature*, **400**, 525–531.
- Watson, A.J., Lovelock, J.E. and Margulis, L. (1978). Methanogenesis, fires and regulation of atmospheric oxygen, *Biosystems*, **10**, 293–298.
- Zeebe, R.E. and Wolf-Gladrow, D. (2001). CO₂ in seawater: equilibrium, kinetics, isotopes. Elsevier Oceanography Series, vol. 65. Elsevier, Amsterdam.

11 The Last Common Ancestor of Modern Cells

David Moreira and Purificación López-García

11.1 The Last Common Ancestor, the Cenancestor, LUCA: What's in a Name?

All living beings share a number of essential features pertaining to their biochemistry and fundamental processes. Some of these features are so complex (an outstanding example is the genetic code) that their probability to have appeared several times independently is almost negligible. This authorizes the concept of a common ancestor to all living beings that possessed these traits, and from which diversification occurred, leading to the emergence of the three domains of life that we recognize today, Archaea, Bacteria, and Eucarya. Nevertheless, the concept of a common ancestor was born well before these universal features were even identified, mostly on the basis of purely philosophical and theoretical considerations.

11.1.1 Some Historical Grounds

The idea that all living beings are linked by a natural process has influenced the scientific thought for a long time. For example in his 1686 *Discourse on Metaphysics*, the German philosopher Leibniz explicitly states the existence of intermediate species between those that are distant, implying that all living beings are related in a great chain, which is intrinsic to the harmony of the universe. Following this inspiration, many scientists have tried to identify the source of diversity and the nature of the process unifying the different species. Darwin's concept of descent with modification and natural selection provided the solid grounds to explain the patterns of similarity and difference among species that have served as a general framework for biology until present days. In *The Origin of Species*, Darwin concluded that a logical outcome of the premises of descent with modification and natural selection is that probably all the organic beings, which have ever lived on this earth have descended from some one primordial form, into which life was first breathed (Darwin 1859). The concept of a common ancestor for all life forms was born.

This idea received strong support during the last century, thanks to the spectacular development of biochemistry and, especially, of molecular biology. Bacteria, eukaryotes and the recently discovered archaea (Woese and Fox 1977a)

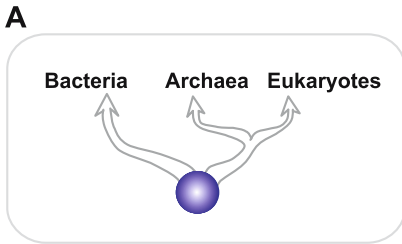
share the basic structural constituents of the cell and the most fundamental metabolic reactions. All life is based on common biochemical themes (Kornberg 2000; Pace 2001).

11.1.2 The Hypothesis of a Cenancestor

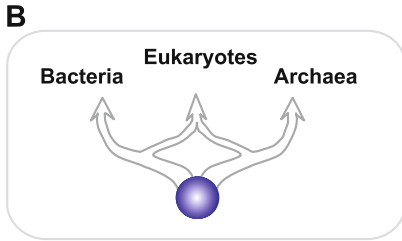
This hypothetical ancestral organism to all living beings has been baptized with different names according to different authors. The most popular are the last common ancestor, the cenancestor (from the Greek *kainos* meaning recent and *koinos* meaning common) (Fitch and Upper 1987) and the last universal common ancestor or LUCA (Forterre and Philippe 1999). Although all evolutionary biologists agree that some type of cenancestor existed, its nature is a matter of intense and highly speculative debates. Most authors favor the idea that the cenancestor was a single organism, an individual cell that existed at a given time and that possessed most of the features (and genes encoding them) that are common to all contemporary organisms. From this single ancestor, the different domains of life would have diverged (Fig 11.1A–D). Others, on the contrary, envisage a population of cells that, as a whole, possessed all those genes, although no single individual did (Kandler 1994; Woese 1998; Woese 2000). This implies that the level of gene exchange and spreading in this population was very high. At some point, however, a particular successful combination of genes occurred in a subpopulation that became isolated and gave rise to a whole line of descent. Kandler, for instance, proposed in his pre-cell theory that bacteria, archaea, and eukaryotes emerged sequentially in this way (Fig 11.1E) (Kandler 1994; Wächtershäuser 2003).

Single cell or population, all researchers agree that the cenancestor was already quite complex, having evolved from simpler entities, and that there was a more or less long evolutionary path from the origin of life to the cenancestor stage. Both, the origin of life and the nature of the cenancestor are different evolutionary questions. Nonetheless, despite the general agreement that the cenancestor was already quite complex, the level of complexity attributed to it varies depending on the model. For Woese, the cenancestor was a relatively primitive entity, which he called a “progenote”, that had not completely evolved the link between genotype and phenotype (Woese and Fox 1977b). For others, the progenote state occurred prior to the cenancestor, which was nearly a modern cell (see review in Doolittle 2000a).

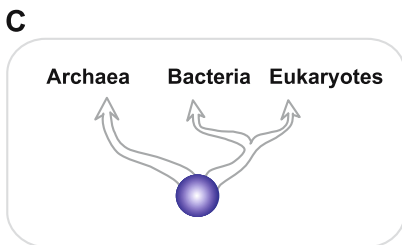
Given their universality, a ribosome-based protein synthesis (translation), a well-developed transcription machinery for the synthesis of structural and messenger RNAs, and an energy-obtaining process based on the generation of a proton gradient across membranes (Gogarten et al. 1989), were among the features that the cenancestor certainly possessed. Other cenancestor properties are, however, much more controversial, such as the existence of a DNA-based genome or even the possession of lipid-based membranes. The occurrence in the cenancestor of these properties was initially deduced mostly from biochemistry and molecular



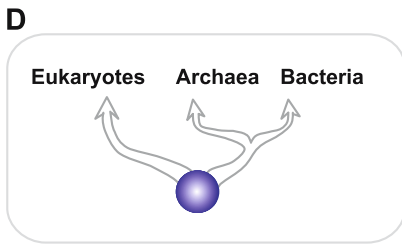
"Bacterial rooting": bacteria and a line leading to Archaea+Eukaryotes emerged from the cencestor. The most popular model.



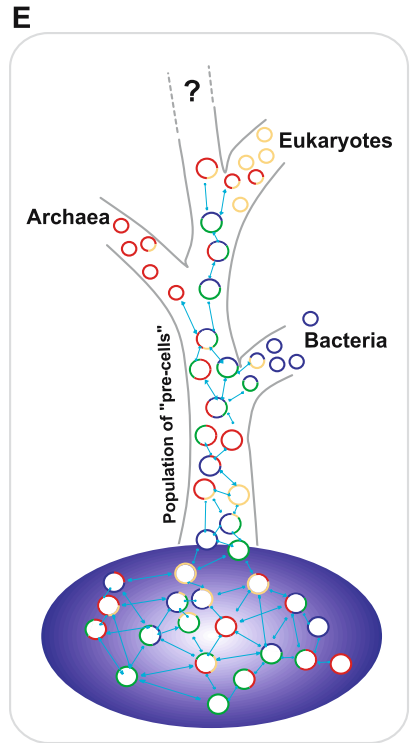
Chimeric models for the origin of eukaryotes. Only the two prokaryotic lineages diverged directly from the cencestor. Eukaryotes would be descendants of archaeal-bacterial symbioses.



"Archaeal rooting": archaea and a line leading to Bacteria+Eukaryotes emerged from the cencestor. This model is the less accredited



"Eukaryotic rooting": eukaryotes and a line leading to prokaryotes emerged from the cencestor. The cencestor might have had eukaryotic characteristics.



"Pre-cell model", by O. Kandler. The cencestor was a multiphenotypical population of pre-cells, characterised by a frequent mutual genetic exchange

Fig. 11.1. Current hypotheses for the evolution of the three domains of life from a last common ancestor or cencestor. The cenancestor stage is indicated by a *blue sphere*. Models **A** to **D** envisage that all the properties attributed to the cenancestor co-existed in the same cell, whereas in **E**, they were present collectively in a population of primitive cells

biology studies. Obviously, many genes (and proteins) were involved in these ancestral machineries and processes. At present, the most powerful tools to infer in more detail which were the genes and proteins already present in the cenancestor are comparative genomics and molecular phylogeny. This type of analysis is greatly facilitated by the increasing number of complete genome sequences available for very different organisms. Although several problems, in particular horizontal gene transfer and differential gene loss, make the reconstruction of the ancestral gene content troublesome (Koonin 2003). We briefly discuss these aspects in the following sections.

11.2 How Did the Cenancestor Make Proteins?

Prokaryotic (archaeal and bacterial) species contain approximately between 500 and 10 000 genes, whereas eukaryotic species contain approximately between 2000 and 30 000 genes. Nevertheless, when the gene content of all available genomes is compared, only ~ 60 genes are found to be common to all of them. This set of genes is almost entirely integrated by genes encoding ribosomal RNA and ribosomal proteins as well as other proteins involved in translation (especially aminoacyl-tRNA synthetases and translation factors) (Koonin 2003). This implies that these genes are ancestral, and provide strong evidence that the cenancestor possessed a fully developed ribosomal-based translation machinery for protein synthesis that was comparable to the one found in modern organisms. Protein synthesis by ribosomes is, therefore, the most universally conserved process. Furthermore, the level of conservation appears so high that the process of protein synthesis has remained practically unchanged for, likely, more than three billion years.

A few of the ~ 60 genes that conform the universal core encode RNA polymerase subunits. RNA polymerase is responsible for the synthesis of messenger and other RNAs from genes (DNA templates) (Koonin 2003). As in the case of translation, this implies that the cenancestor possessed at least part of the transcription machinery found in contemporary organisms. Nevertheless, the degree of conservation of the transcription machinery is not as high, since several RNA polymerase subunits and transcription factors are not universally distributed.

Transcription and translation are among the most conserved processes in cells, and can be traced back to the cenancestor. The fact that both processes depend primarily on structural and catalytic RNA molecules, e.g., the ribosomal RNAs, indicates that RNA played an essential role since very early in cell evolution. This and the recent discovery of ribozymes (small catalytic RNAs) have been important elements leading to the proposal of an “RNA!world” (for review, see Joyce 2002). According to this model, there was a very early evolutionary stage when RNA carried both, information-storing and catalytic functions, which are performed today by DNA and proteins, respectively. While being broadly accepted, this model remains, however, purely hypothetical.

11.3 What Was the Nature of the Genetic Material?

Despite the fact that, in all contemporary cells, DNA is the molecule where genetic information is stored, only three out of the ~ 60 universal genes are related to DNA replication and/or repair. These are one DNA polymerase subunit, one exonuclease and one topoisomerase (Koonin 2003). From all the genes known to be involved in DNA replication in organisms of the three domains of life, many are shared by archaea and eukaryotes but are absent from bacteria. The latter apparently possesses completely unrelated genes encoding the proteins that perform the equivalent functions. Various hypotheses have been proposed to explain the profound differences between the bacterial and archaeal/eukaryotic replication machineries. One of them even postulates that the cenancestor did not possess a DNA genome at all. According to this model, the cenancestor would have had an RNA genome, and DNA replication would have evolved twice independently, once in the bacterial line of descent and other in the lineage leading to archaea and eukaryotes (Forterre 2002; Leipe et al. 1999). However, this model is contested by different lines of evidence that, put together, suggest that the cenancestor already possessed a DNA genome. First, although few, universally conserved proteins and protein domains involved in DNA metabolism indeed exist (Giraldo 2003). Second, RNA is much more error-prone than DNA due to its higher mutation rate, so that single RNA molecules cannot exceed a certain size (Eigen limit) without falling into replicative catastrophe (Eigen 1971, 2002). This limit is so small, $\sim 30\text{--}50\text{kb}$, that an RNA molecule could contain only a few dozen genes. Recent estimates suggest that the, already quite complex, cenancestor may have had probably more than 600 genes (Koonin 2003). If its genome was made of RNA, many RNA molecules would have been required to contain all those genes, which poses a serious problem of stability during replication and partition among daughter cells. The existence of this problem is attested by the characteristics of DNA and RNA viruses today. Whereas DNA viruses can have genomes that reach very large sizes, up to $\sim 1\text{Mbp}$ (Raoult et al. 2004), RNA viruses' genomes do not exceed $\sim 30\text{kb}$ (Domingo and Holland 1997).

The remaining models propose that the cenancestor had a DNA-based genome. One of these explains the dichotomy of DNA replication in bacteria and archaea/eukaryotes by stating that, whereas transcription and translation were already well developed in the cenancestor, DNA replication was still very primitive. DNA replication would have been improved and refined after, or simultaneously to, the separation of the two lines of descent leading to the bacteria and to the archaea/eukaryotes (Olsen and Woese 1997). A contrasting model could be that the cenancestor had a highly complex DNA-based metabolism and contained the ancestral versions of the proteins found today in both the bacterial and the archaeal/eukaryotic replication machineries. One set of proteins would be involved in replication, whereas the other would be specialized in DNA repair. During the speciation of the two lines, only one set of proteins would have been retained in each line of descent. Another proposal suggests that the replication machinery was already well developed in the cenancestor, but that it evolved

very fast in one or the two lineages descending from the cenancestor to the point that the similarity between the homologous genes in both lines is no longer recognizable (Moreira 2000; Olsen and Woese 1997). Finally, an additional hypothesis suggests that archaea and eukaryotes have retained the ancestral DNA replication system, whereas the bacteria have seen their machinery replaced by genes imported from viruses (Forterre 1999). However, there is no evidence for this hypothesis but, on the contrary, many examples of the opposite, i.e., numerous acquisitions of replication-related genes by viruses from the genome of their cellular hosts (Moreira 2000). In conclusion, although still a matter of debate, it appears likely that the cenancestor had a DNA genome. However, the mechanism for its replication and the early divergence of the replication machineries in the bacteria and the archaea/eukaryotes remain a mystery.

11.4 What Did the Cellular Metabolism Look Like?

The question of how metabolism looked in the cenancestor is a difficult and complex one, and hence, is generally put aside in descriptions of a putative ancestor. The theoretical proposal of a cenancestor is fundamentally based on conserved genes related to the storage, expression and transmission of the genetic information present in all contemporary organisms (for recent reviews, see Doolittle 2000a, 2000b; Koonin 2003). However, genes involved in energy and carbon metabolism not only display a patchy distribution in organisms of the three domains of life, but they very often belong to large multigenic families whose members have been recruited for different functions in various metabolic pathways. Furthermore, horizontal gene transfer is known to affect preferentially metabolic genes, since they may confer an immediate adaptive advantage. Therefore, the reconstruction of ancestral metabolic pathways is frequently masked by a complex history of gene duplication and differential enzyme recruitment (Castresana and Moreira 1999).

Discussions and controversies about early metabolism do take place, but they concern the very first metabolism at the time when life arose rather than the metabolic traits in a more evolved cenancestor. These discussions are model-dependent. If the first living beings were heterotrophs feeding upon an organic prebiotic soup as proposed classically (Oparin 1938), then fermentation, which is mechanistically simple, may have been the first way of gaining energy (Broda 1970), while cell building blocks were directly uptaken from the soup. If the first living beings were chemolithoautotrophs initially developed on pyritic (FeS_2) surfaces (Wächtershäuser 1988) or in iron monosulfide (FeS) tridimensional compartments (Russell and Hall 1997), energy was derived from redox reactions involving inorganic molecules such as H_2S , H_2 and FeS . In this case, organic molecules were synthesized *de novo* from CO_2 or CO by an ancestral, yet undetermined, metabolic pathway. Today, four different pathways of autotrophic carbon fixation are known (revised in Peretó et al. 1999). The Calvin–Benson

cycle (reductive pentose-phosphate pathway) is found in oxygenic photosynthesizers such as cyanobacteria and plants, but also in other autotrophic prokaryotes. The Arnon cycle (reductive citric acid pathway) is found in several bacteria and archaea. The Wood–Ljungdahl cycle (reductive acetyl-CoA pathway) is found in acetogenic and sulfate-reducing bacteria and methanogenic archaea. Finally, the hydroxypropionate pathway is found in the green non-sulfur bacterium *Chloroflexus* sp. and a few thermophilic archaea. There is general agreement to consider that the Calvin–Benson cycle appeared relatively late during bacterial evolution. However, which of the other three is older is a matter of debate. The Arnon cycle is proposed to be older by Wächtershäuser and other adherents of the pyrite-based chemoautotrophic origin of life because of its wide distribution in bacteria and archaea (Wächtershäuser 1990). Nevertheless, a reductive acetyl-CoA pathway is proposed to be the primordial autotrophic pathway because of its higher simplicity (acetate, a two-carbon molecule is synthesized) and its exergonic (energy-releasing) nature under certain hydrothermal conditions (Russell and Martin 2004). Finally, although less known, the hydroxypropionate pathway is also very simple and has been proposed to be the first autotrophic pathway at least in phototrophic bacteria (Peretó et al. 1999).

Although the metabolic properties of the cenancestor are not generally discussed, if any or several of these pathways had developed in earlier times, the cenancestor must have possessed them. In any case, the presence of a universal highly conserved membrane-bound ATPase indicates that the cenancestor was able to produce energy in the form of ATP by generating a proton gradient across the cell membrane. What is unclear is the type of electron donors and acceptors required to generate this proton gradient, although it might have been able to use a variety of oxidized inorganic molecules as electron acceptors. It is also likely that the cenancestor was able to carry out a simple heterotrophic metabolism, at least some type of fermentation, but whether it was a full autotroph or not remains an open question.

11.5 Was the Cenancestor Membrane-Bounded?

All contemporary cells are surrounded by a plasma membrane that is made out of phospholipids generally organized in bilayers. However, as in the case of replication, there exist profound differences between, this time, the membrane lipids of archaea and the membrane lipids of bacteria and eukaryotes. In archaea, phospholipids are made out of generally isoprenoid lateral chains that are bounded by ether linkages to glycerol-1-phosphate, whereas in bacteria and eukaryotes phospholipids are made out of fatty acids bounded by ester linkages to glycerol-3-phosphate. From these differences, the most fundamental is the opposite stereochemistry of the glycerol-phosphate, since some bacteria and eukaryotes make ether linkages under certain circumstances, and since archaea do make fatty acid ether phospholipids as well (Peretó et al. 2004). The

enzymes responsible for the synthesis of glycerol-1-phosphate and glycerol-3-phosphate are not homologous in archaea and bacteria/eukaryotes, but belong to two different enzymatic families. This difference is so hard to explain that some authors have proposed that the cenancestor had not yet reached a cellular stage. Membrane lipids, and cells simultaneously, would have evolved independently from a non-enzymatic lipid synthesis pathway to generate bacteria and archaea (Koga et al. 1998). Other authors propose that the cenancestor did not possess a lipid membrane at all, and that membrane lipids (and their biosynthetic pathways) were invented twice, in each one of the lineages that diverged from the cenancestor leading to the bacteria and to the archaea. Instead of lipids, the cenancestor would have been endowed with a mineral membrane made out of iron monosulfide; cells would be mineral compartments in a particular kind of hydrothermal chimneys (Martin and Russell 2003). Finally, another, less radical option, is that the cenancestor had a lipid membrane, but that it was heterochiral, i.e., it was composed by a mixture of lipids built upon glycerol-1-phosphate and glycerol-3-phosphate (Wächtershäuser 2003). A subsequent specialization of the biosynthetic pathways to yield the two types of homochiral membranes would have accompanied the speciation of archaea and bacteria. Recent phylogenetic analyses of the enzymes involved in the synthesis of glycerol phosphate strongly suggest that the cenancestor possessed a non-stereospecific pathway of phospholipid biosynthesis and was therefore endowed with heterochiral membranes (Peretó et al. 2004). A cenancestor endowed with a lipid membrane would be in agreement with the occurrence of several membrane-bound proteins extremely well conserved, notably the proton-pump ATPases (Gogarten et al. 1989) and the signal recognition particle, SRP (Gribaldo and Cammarano 1998).

11.6 Other Unresolved Questions

As we have seen above, some properties of the hypothetical cenancestor appear more or less well-defined, such as the possession of a quite modern transcription and translation machinery and, most likely, the existence of phospholipid membranes. Other features remain obscure, such as the type of carbon and energy metabolism or the replication of the genetic material. However, there are additional questions that remain open and that have been the subject of lively debates. Among these, whether the cenancestor was hyperthermophilic or not, and whether it was *simple* or *complex*.

The proposal of a hyperthermophilic cenancestor is linked to the discovery of hyperthermophilic bacteria and archaea growing optimally at $>80^{\circ}\text{C}$, and to the first proposals of a hot, autotrophic origin of life in a warmer early Earth with extended hydrothermal activity (Achenbach-Richter et al. 1987; Pace 1991). The most important argument used was that hyperthermophilic prokaryotes branched at the most basal positions in phylogenetic trees (Stetter 1996). The first criticisms to a hot origin of life derived from the fact that RNA and

other important biomolecules have relatively short life-times at high temperatures (Lazcano and Miller 1996). However, as mentioned before, the origin of life and the cenancestor were separated in time, and might have occurred in different environmental conditions (Arrhenius et al. 1999). Thus, some authors propose that the cenancestor was hyperthermophilic, and that the origin of life might have taken place at much lower temperatures, but only hyperthermophiles could survive the late heavy meteorite bombardment ~ 3.9 Ga ago (Gogarten-Boekels et al. 1995). However, the most important criticism to a hypothetical hyperthermophilic ancestor derives from recent phylogenetic analyses. On the one hand, computer reconstruction of ancestral ribosomal RNA sequences suggests that the content of guanine and cytosine in the cenancestor's RNA was incompatible with life at $>80^\circ\text{C}$ (Galtier et al. 1999). However, the analysis of the same datasets using other methods questions this conclusion favoring, on the contrary, a hyperthermophilic cenancestor (Di Giulio 2000). On the other hand, refined phylogenetic analyses of ribosomal RNA sequences suggest that the basal emergence of hyperthermophilic bacteria was an artifact of phylogenetic tree reconstruction and favor the idea that they adapted secondarily to hyperthermophily (Brochier and Philippe 2002). However, as in the previous case, the use of other phylogenetic methods comes up with opposite results (Di Giulio 2003). At any rate, although the situation in bacteria is highly controversial, there is general agreement that the ancestor of the archaea was indeed hyperthermophilic (Forterre et al. 2002). Therefore, if the bacterial ancestor was also a hyperthermophile, then the most parsimonious conclusion is that the cenancestor was hyperthermophilic. If the bacterial ancestor was not hyperthermophilic, then it will be very difficult to infer the type of environmental conditions in which the cenancestor thrived. At any rate, all the current analyses and proposals appear compatible with the occurrence of a thermophilic ($60\text{--}80^\circ\text{C}$) cenancestor (López-García 1999).

Another controversial issue concerns the level of complexity that the cenancestor possessed. As stated above, it is clear that it was already very complex. Even the authors that call it simple, envisage that its genome contained several hundred genes ($\sim 600\text{--}1000$), which is in the range of the simplest present-day prokaryotes (Koonin 2003). Most authors imagine an ancestor that was also *structurally simple*, i.e., with a cellular organization resembling that of today's prokaryotes. An ancestor of this type would have the genetic material directly immersed in the cytoplasm, where the replication, transcription and translation would take place. This type of ancestor is supported by the widely accepted bacterial rooting of the tree of life, i.e., that the root of the tree of life lies between the bacteria and the archaea/eukaryotes (Fig. 11.1A) (Woese et al. 1990), but would be also compatible with two alternative tree topologies (Fig. 11.1B,C). However, this has been challenged by authors that propose a eukaryotic rooting of the tree of life, i.e., that the root lies in between the eukaryotes and a branch leading to the two prokaryotic

domains (Fig. 11.1D) (Brinkmann and Philippe 1999; Philippe and Forterre 1999). This eukaryotic rooting would still be compatible with a prokaryote-like cenancestor, but it opens the possibility that the cenancestor had some of the traits that characterize modern eukaryotes, in particular the presence of a membrane-bound nucleus and of many small RNA molecules claimed to be relics of a hypothetical RNA world (Poole et al. 1999). In this model, prokaryotes would be derived by a reductive process from more complex eukaryotic-like ancestors. However, although the position of the root is indeed an open question, models proposing a eukaryotic-like cenancestor do not explain how such a complex entity was built from the prebiotic world. In this sense, a simpler, prokaryotic-like cenancestor appears much more parsimonious in evolutionary terms.

11.7 Perspectives

Obviously many questions have to be answered before achieving an accurate picture of the putative cenancestor. Fortunately, it seems likely that a number of these questions can be addressed thanks to the increasing amount of data derived from comparative genomics. Nonetheless, a crucial issue will be to determine the impact of horizontal gene transfer in evolution because, if horizontal gene transfer has been rampant all along life history as some authors suggest (Doolittle 2000b), it would then be very difficult or even impossible to reconstruct any ancestor. In a scenario of frequent or even massive horizontal gene transfer, a single cenancestor that contained all the genes ancestral to those shared among the three domains of life did not likely exist. Rather, these ancestral genes were probably present in different organisms and at different times (Zhaxybayeva and Gogarten 2004). The application of population genetics methods to the study of gene emergence and extinction over long timescales may shed some light into this problem.

Another open question concerns the origin and evolution of viruses, and how these have affected cellular evolution. For instance, they may have contributed significantly to horizontal gene transfer serving as vehicles of gene exchange. They may have also helped to increase the evolutionary rate of many genes thus contributing to the development of novel gene functions.

An alternative source of information to reconstruct a model portrait of the cenancestor may come from the generation of more accurate models of the early Earth. This could help to delineate the environmental conditions where life arose and first evolved, giving clues, for instance, as to the putative hyperthermophilic nature of the cenancestor, and/or to the most likely metabolic pathways it utilized. Similarly, resolving the controversies that currently exist on the earliest microfossils and increasing the number of unambiguous fossil data will be most helpful to establish a likely chronology of early life evolution.

References

- Achenbach-Richter, L., R. Gupta, K.O. Stetter, and C.R. Woese. 1987. Were the original Eubacteria thermophiles? *Syst. Appl. Microbiol.* 9: 34–39.
- Arrhenius, G., J.L. Bada, G.F. Joyce, A. Lazcano, S. Miller, and L.E. Orgel. 1999. Origin and ancestor: separate environments. *Science* 283: 792.
- Brinkmann, H., and H. Philippe. 1999. Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Mol. Biol. Evol.* 16: 817–825.
- Brochier, C., and H. Philippe. 2002. Phylogeny: A non-hyperthermophilic ancestor for Bacteria. *Nature* 417: 244.
- Broda, E. 1970. The evolution of bioenergetic processes. *Prog. Biophys. Mol. Biol.* 21: 143–208.
- Castresana, J., and D. Moreira. 1999. Respiratory chains in the last common ancestor of living organisms. *J. Mol. Evol.* 49: 453–460.
- Darwin, C. 1859. *The Origin of Species by Means of Natural Selection*. J Murray, London.
- Di Giulio, M. 2000. The universal ancestor lived in a thermophilic or hyperthermophilic environment. *J. Theor. Biol.* 203: 203–213.
- Di Giulio, M. 2003. The ancestor of the Bacteria domain was a hyperthermophile. *J. Theor. Biol.* 224: 277–283.
- Domingo, E., and J.J. Holland. 1997. RNA virus mutations and fitness for survival. *Ann. Rev. Microbiol.* 51: 151–178.
- Doolittle, R.F. 2000a. Searching for the common ancestor. *Res. Microbiol.* 151: 85–89.
- Doolittle, W.F. 2000b. The nature of the universal ancestor and the evolution of the proteome. *Curr. Opin. Struct. Biol.* 10: 355–358.
- Eigen, M. 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58: 465–523.
- Eigen, M. 2002. Error catastrophe and antiviral strategy. *Proc. Natl. Acad. Sci. USA* 99: 13374–13376.
- Fitch, W.M., and K. Upper. 1987. The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* 52: 759–767.
- Forterre, P. 1999. Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. *Mol. Microbiol.* 33: 457–465.
- Forterre, P. 2002. The origin of DNA genomes and DNA replication proteins. *Curr. Opin. Microbiol.* 5: 525–532.
- Forterre, P., C. Brochier, and H. Philippe. 2002. Evolution of the Archaea. *Theor. Pop. Biol.* 61: 409–422.
- Forterre, P., and H. Philippe. 1999. The last universal common ancestor (LUCA), simple or complex? *Biol. Bull.* 196: 373–375.
- Galtier, N., N. Tourasse, and M. Gouy. 1999. A nonhyperthermophilic common ancestor to extant life forms. *Sciences* 283: 220–221.
- Giraldo, R. 2003. Common domains in the initiators of DNA replication in Bacteria, Archaea and Eukarya: combined structural, functional and phylogenetic perspectives. *FEMS Microbiol. Rev.* 26: 533–554.

- Gogarten, J.P., H. Kibak, P. Dittrich, L. Taiz, E.J. Bowman, B.J. Bowman, M.F. Manolson, R.J. Poole, T. Date, T. Oshima, J. Konishi, K. Denda, and M. Yoshida. 1989. Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proc. Nat. Acad. Sci. USA* 86: 6661–6665.
- Gogarten-Boekels, M., E. Hilario, and J.P. Gogarten. 1995. The effects of heavy meteorite bombardment on the early evolution—the emergence of the three domains of life. *Orig. Life Evol. Biosph.* 25: 251–264.
- Gribaldo, S., and P. Cammarano. 1998. The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. *J. Mol. Evol.* 47: 508–516.
- Joyce, G.F. 2002. The antiquity of RNA-based evolution. *Nature* 418: 214–221.
- Kandler, O. 1994. The early diversification of life. In: Bengtson, S. (ed.) *Early Life on Earth*, pp. 152–160, Columbia University Press, New York.
- Koga, Y., T. Kyuragi, M. Nishihara, and N. Sone. 1998. Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. *J. Mol. Evol.* 46: 54–63.
- Koonin, E.V. 2003. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* 1: 127–136.
- Kornberg, A. 2000. Ten commandments: lessons from the enzymology of DNA replication. *J. Bacteriol.* 182: 3613–3618.
- Lazcano, A., and S.L. Miller. 1996. The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. *Cell* 85: 793–798.
- Leipe, D.D., L. Aravind, and E.V. Koonin. 1999. Did DNA replication evolve twice independently? *Nucleic Acids Res.* 27: 3389–3401.
- López-García, P. 1999. DNA supercoiling and temperature adaptation: a clue to early diversification of life? *J. Mol. Evol.* 49: 439–452.
- Martin, W., and M.J. Russell. 2003. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 358: 59–83.
- Moreira, D. 2000. Multiple independent horizontal transfers of informational genes from bacteria to plasmids and phages: implications for the origin of bacterial replication machinery. *Mol. Microbiol.* 35: 1–5.
- Olsen, G.J., and C.R. Woese. 1997. Archaeal genomics: an overview. *Cell* 89: 991–994.
- Oparin, A.I. 1938. *The Origin of Life*. Mac Millan, New York.
- Pace, N.R. 1991. Origin of life—facing up to the physical setting. *Cell* 65: 531–533.
- Pace, N.R. 2001. The universal nature of biochemistry. *Proc. Natl. Acad. Sci. USA* 98: 805–808.
- Peretó, J., P. López-García, and D. Moreira. 2004. Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem. Sci.* 29: 469–477.
- Peretó, J.G., A.M. Velasco, A. Becerra, and A. Lazcano. 1999. Comparative biochemistry of CO₂ fixation and the evolution of autotrophy. *Int. Microbiol.* 2: 3–10.
- Philippe, H., and P. Forterre. 1999. The rooting of the universal tree of life is not reliable. *J. Mol. Evol.* 49: 509–523.
- Poole, A., D. Jeffares, and D. Penny. 1999. Early evolution: prokaryotes, the new kids on the block. *Bioessays* 21: 880–889.

- Raoult, D., S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan, and J.M. Claverie. 2004. The 1.2-megabase genome sequence of Mimivirus. *Science* 306: 1344–1350.
- Russell, M., and A.J. Hall. 1997. The emergence of life from iron monosulfide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc. Lond.* 154: 377–402.
- Russell, M.J., and W. Martin. 2004. The rocky roots of the acetyl-CoA pathway. *Trends Biochem. Sci.* 29: 358–363.
- Stetter, K.O. 1996. Hyperthermophilic prokaryotes. *FEMS Microbiol. Rev.* 18: 149–158.
- Wächtershäuser, G. 1990. Evolution of the first metabolic cycles. *Proc. Natl. Acad. Sci. USA* 87: 200–204.
- Wächtershäuser, G. 2003. From pre-cells to Eukarya—a tale of two lipids. *Mol. Microbiol.* 47: 13–22.
- Wächtershäuser, G. 1988. Before enzymes and templates: theory of surface metabolism. *Microbiol. Rev.* 52: 452–484.
- Woese, C. 1998. The universal ancestor. *Proc. Natl. Acad. Sci. USA* 95: 6854–6859.
- Woese, C.R. 2000. Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci. USA* 97: 8392–8396.
- Woese, C.R., and G.E. Fox. 1977a. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* 74: 5088–5090.
- Woese, C.R., and G.E. Fox. 1977b. The concept of cellular evolution. *J. Mol. Evol.* 10: 1–6.
- Woese, C.R., O. Kandler, and M.L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* 87: 4576–4579.
- Zhaxybayeva, O., and J.P. Gogarten. 2004. Cladogenesis, coalescence and the evolution of the three domains of life. *Trends Genet.* 20: 182–187.

12 An Extreme Environment on Earth: Deep-Sea Hydrothermal Vents. Lessons for Exploration of Mars and Europa

Daniel Prieur

Observed from space, Earth clearly deserves the name of *Blue Planet*: within the solar system, it is the only body with liquid water at the surface. The oceans cover 70% of the Earth's surface and have an average depth of 3800m (Jan-nash and Taylor 1984). With the exception of changing colors, temperature or waving, this huge amount of water may seem homogeneous. Nevertheless, the oceanic domain shows a great diversity. Particularly, deep-sea hydrothermal vents show features totally unexpected until their discovery less than 30 years ago. Such extreme environments, supporting a wide diversity of living organisms were not thought to exist on Earth. For some of the first scientists involved in their explorations, they might represent a reminiscence of the primitive Earth, when life first arose. In this chapter, we will successively examine the invertebrate and microbial communities associated with deep-sea hydrothermal vents, and for the latest, some of the most fascinating adaptive mechanisms to extreme conditions, not so far from those suspected on Mars or Europa.

12.1 Some Features of Oceanic Environment

All organisms living on Earth are influenced by various physicochemical parameters, and in particular by temperature, light intensity, pH, concentration of salts, metals, organic and inorganic compounds, oxygen levels, water activity, and atmospheric or hydrostatic pressure. Each combination of these parameters is characteristic of a biotope: a geographical entity with a certain area/volume, in which individual members of specific living communities carry out their life cycles.

According to the ocean or sea concerned, all these parameters display various values (Prieur 1992). For instance:

- Temperature may vary from -2°C in polar waters (salty water does not freeze) to 35°C for surface waters of the Persian Gulf.
- The average salinity is about 35‰, but may reach 41‰ for areas exposed to severe evaporation (Red Sea). Moreover, salt evaporating ponds are saturated, but rivers bring fresh water into the seas, and a salinity gradient occurs in estuaries.

- Organic matter may fluctuate greatly with very rich (eutrophic) and poor (oligotrophic) zones.
- Also oxygen concentrations vary from saturation at the surface to complete anoxia in some deep water layers or within the sediments.
- Hydrostatic pressure increases according to depth (10 bars or 1MPa for 10 meters), with a maximum of 110MPa for the deepest oceanic trenches.

Finally, total darkness exists beyond a few hundred meters depth. The lack of light prevents photosynthesis, and consequently primary production (organic matter synthesis by autotrophic plants), but also the release of molecular oxygen into the environment. Deep oceanic environments are cold (average temperature: 2°C), exposed to elevated hydrostatic pressure, and from a trophic point of view, depend on the ocean upper layers: only corpses or feces from organisms living in the photic zones, or what remains after in route bacterial degradation sediment to the sea floor (Jannash and Taylor 1984). Similarly, molecular oxygen present in the deep water layers is a product of photosynthesis occurring in surface waters. Until 1977, ocean bottoms were known as ecosystems with low animal density.

The definition of a biotope as an extreme environment is certainly influenced by our own perception of the environment. However, any biotope characterized by a very low or very high value for one of the main parameters influencing life cycles may be qualified as extreme. With this respect, the deep-sea, dark, oligotrophic, cold and pressurized, is one of these extreme environments, where nevertheless bacteria, invertebrates and fish thrive.

12.2 Deep-Sea Hydrothermal Vents

The discovery of deep-sea hydrothermal vents in 1977 (Corliss et al. 1979) strongly modified our knowledge of deep oceanic environments. Hydrothermalism is a consequence of plate tectonics. Most mid-ocean plates are delimited by submarine mountain chains, called oceanic ridges, whose lengths reach 60 000 km. These plates separate and are intensively faulted; consequently cold seawater penetrates into the fractures of the oceanic crust. During its descent into the crust, water progressively warms up, thus creating a convective water circulation into the oceanic crust. This water does not only exchange energy with basalts but also it chemically interacts with it. This is well exemplified by the black or white smokers, which correspond to the escape of warm (>300°C) hydrothermal water rich in sulfurs of Fe, Mn, Ni, Cr, etc., and also in methane dissolved from the basalts.

A series of dives was organized in 1977 in the Galapagos area at a depth of 2000m, using the manned deep-sea submersible “Alvin”. Alvin passengers and pilot were strongly surprised: around warm water springs, they watched for the first time unexpected, beautiful and luxuriant animal communities (Corliss et al. 1979). Two years later, in 1979, again on the East Pacific

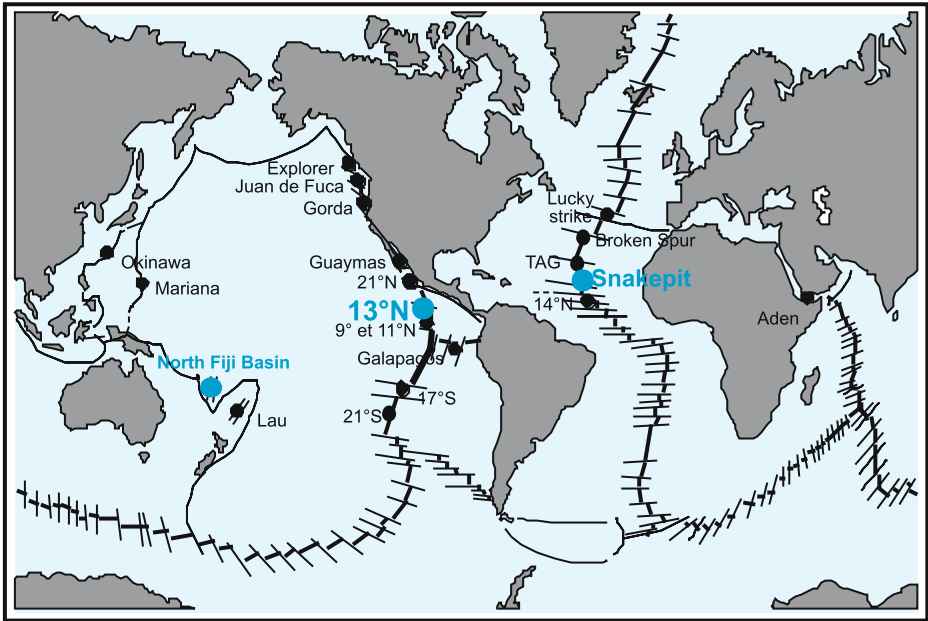


Fig. 12.1. Locations of deep-sea hydrothermal vents (courtesy of Yves Fouquet, IFREMER, Centre de Brest, France)

Rise, another oceanic expedition discovered the now famous “black smokers” (Fig. 12.2), spectacular mineral constructions from which particle-rich fluids vent out at speeds up to 2m s^{-1} , and temperatures up to 350°C (water remains liquid as a consequence of elevated hydrostatic pressure existing at these depths). Probably one of the most extreme environments on Earth was discovered.

Hydrothermal vents have now been discovered in several tectonically active zones of the sea floor, in ridges and back-arc basins located in the Pacific and Atlantic Oceans at depths ranging from 800 to 3600m (Fig. 12.1). After a complex circulation, hydrothermal fluids are vented out into cold and oxygenated deep water. These hydrothermal fluids are acidic, reduced and enriched in heavy metals (iron, manganese, zinc, copper, etc.) and inorganic molecules such as hydrogen, methane, hydrogen sulfide, carbon dioxide, carbon monoxide (Table 12.1). When these fluids mix with descending cold seawater before venting out, their temperature decreases (10 to 40°C). This is where the invertebrate communities thrive. In case no mixing has occurred before, when the hot fluids vent out, their dissolved mineral elements precipitate in contact with cold seawater, generating the fascinating structures called black smokers or hydrothermal chimneys, from which thermophilic microorganisms were discovered and described. Microbiologists could not deny themselves the investigation of such ecosystems.

Table 12.1. Chemical composition of standard seawater and hydrothermal fluids from the Mid-Atlantic Ridge (adapted from Charlou et al. (2002), reprinted with permission from Elsevier)

sw	Logatchev (14°45'N)	MARK-1/2 (23°N)	TAG (26°N)	Broken Spur (29°N)	Rainbow (36°14'N)	Lucky Strike (37°17'N)	Menez Gwen (37°50'N)
Depth (m)	3000	3460	3670	3200/3300	2300	1700	850
Temp. (°C)	347/352	335/350	321/290	356-364	365	170/364	275/284
pH	3.3	3.9/3.7	3.1	-	2.8	3.5/3.7	4.2/4.3
Si(OH) ₄ (mM)	8.2	18.2/18.3	22.0	-	6.9	11.5/16.3	7.7/11.6
Cl (mM)	515	559/559	659	469	750	422/533	357/381
Br (μM)	818	847	880/1045	749/765	1178	735/924	666/710
SO ₄ (mM)	0	0	0	0	0	0	0
Na (mM)	430	510/509	584	419/422	553	347/446	312/319
Li (μM)	245	843/849	411	1006/1033	340	278/357	238/274
K (mM)	21.9	23.6/23.9	18.0	18.1/19.6	20.4	21.1/26.7	22.1/23.8
Rb (μM)	27.7	10.5/10.8	10	13.0/13.6	36.9	22.7/39.1	20.3/29.4
Cs (nM)	385	177/181	110	139/146	333	200/280	330
Mg (mM)	53	0	0	0	0	0	0
Ca (mM)	10.2	9.9/10.5	26.0	11.8/12.8	66.6	31.3/38.2	29.7/33.1
Sr (μM)	87	50/51	99	42.9/48.0	200	67/119	100/111
Ba (μM)	0.14	>4.3	>19	>12.9/>21.3	>67	10/52	>12
Fe (μM)	<0.001	2180/1832	1640	1684/2156	24050	30/863	<2/18
Mn (μM)	<0.001	491/493	1000	250/260	2250	84/446	59/71

Table 12.1. (continued)

sw	Logatchev (14°45'N)	MARK-1/2 (23°N)	TAG (26°N)	Broken Spur (29°N)	Rainbow (36°14'N)	Lucky Strike (37°17'N)	Menez Gwen (37°50'N)
Cu (μM)	15/50	17/10	150	28.3/69.6	121/162	4/26	0.6/3
Zn (μM)	25/30	50/47	46	40.8/88.0	115/185	5/57	2.4/4.3
<i>Gases</i>							
H ₂ S (mM)	0.5/0.8	5.9/5.9	6.7	8.5/11.0	1.2	2.5/3.0	<1.5
CO ₂ (mM)	10.1	5.2/6.7	2.9/3.4	6.9/7.1	16	13/28	17/20
δ ¹³ C (CO ₂) (‰ PDB)	-5.1/-5.9	-	-8.4/-10.0	-9.0	-3.15	-7.2/-10.6	-6.8/-9.1
CH ₄ (mM)	2.1	0.023/0.062	0.124/0.147	0.065/0.13	2.5	0.50/0.97	1.35/2.63
δ ¹³ C (CH ₄) (‰ PDB)	-13.6	-	-8.0/-9.5	-18/-19	-15.8	-12.7/-13.7	-18.8/-19.6
CO (nM)	-	-	-	-	5000	-	-
Ar (μM)	12	17/35	20/40	-	-	11/30	11/38
N ₂ (mM)	3.0	1.20/3.37	0.90/0.89	-	1.8	0.61/0.97	0.60/1.90
H ₂ (mM)	12.0	0.19/0.48	0.15/0.37	0.43/1.03	16	0.02/0.73	0.024/0.048

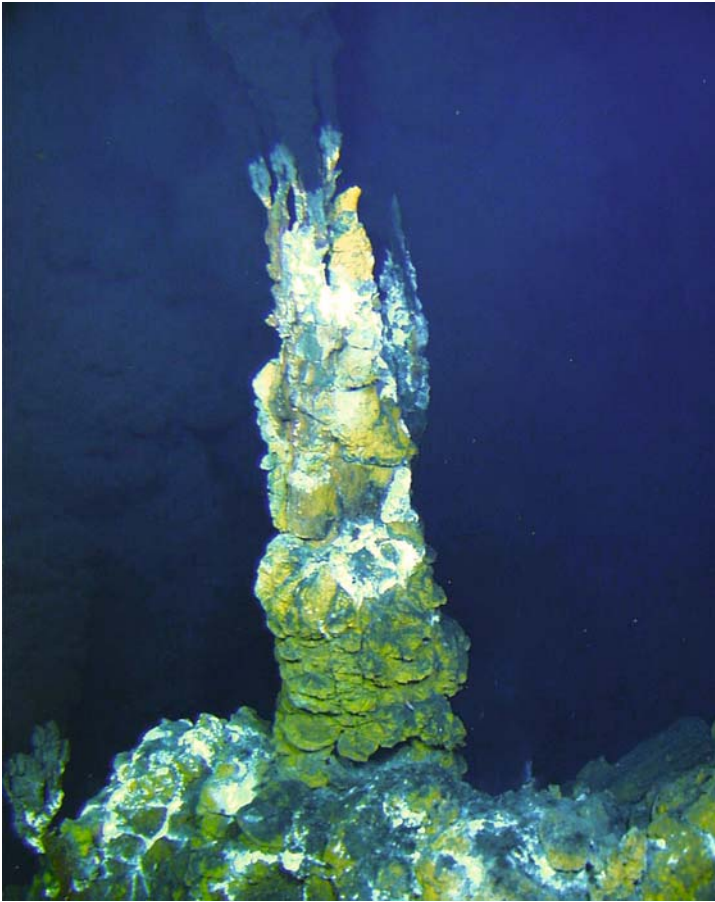


Fig. 12.2. An active black smoker at the East Pacific Rise (courtesy of Daniel Desbruyères, IFREMER, Centre de Brest, France)

12.3 Highly Efficient Symbioses

The discovery of deep-sea hydrothermal vents was a great surprise: luxuriant animal communities thrived in close connection with warm water emissions: tube worms, bivalves, polychaetous annelids, gastropods or crustaceans depending on the geographic area. How does one explain such animal (primary consumers) concentrations, where there is no photosynthetic primary production? Tissues of tube worms (*Vestimentifera*) and bivalves were examined with SEM and TEM and analyzed in order to evidence particular enzymes. These investigations lead to the following conclusive answers: these invertebrates live in close association with intracellular chemolithoautotrophic bacteria that utilize the energy gained from aerobic oxidation of sulfides (in some cases aerobic oxi-

dation of methane) to fix carbon dioxide into organic matter (Cavanaugh et al. 1981).

12.3.1 Vestimentifera

This particular symbiosis is particularly efficient for the Vestimentifera (*Riftia pachyptila*, Fig. 12.3) and other tube worms from the Eastern Pacific. This animal may reach two meters high and lives within a tube made of chitin and proteins out of which a plume of red gills appears. It lacks a digestive tract (no mouth or anus) but has a specific organ called the trophosome, full of autotrophic sulfur-oxidizing bacteria (Cavanaugh et al. 1981; Felbeck 1985). In its environment, the worm collects the nutrients required for bacterial life: carbon dioxide as a carbon source, hydrogen sulfide as energy source and molecular oxygen as electron acceptor. These last two molecules are transported on two distinct sites of a blood pigment rather similar to haemoglobin.

12.3.2 Molluscs

Bivalve molluscs also harbor symbionts. The large bivalve *Calyptogena magnifica*, which lives in eastern Pacific vent sites, has a digestive tract that is most



Fig. 12.3. Tube worms (*Riftia pachyptila*) showing their red gills at the East Pacific Rise (courtesy of Daniel Desbruyères, IFREMER, Centre de Brest, France)

probably not functional (Cavanaugh 1983). The symbionts are again sulfur oxidizers and live in specific cells of the gill filaments (Le Pennec and Prieur 1985). In the case of mussels, such as *Bathymodiolus* (Fig. 12.4), present with different species in all vent sites explored so far, the digestive tract is functional, but symbionts are still present within their gill filaments (Le Pennec and Prieur 1985). Certain *Bathymodiolus* species possess methanotrophic symbionts, and double symbiosis (methanotrophic bacteria and sulfur oxidizing bacteria within the same host) have been described for certain Mid-Atlantic Ridge mussel populations (Cavanaugh et al. 1992). Today, all the symbionts (this is also true for *Vestimentifera* symbionts) cannot be cultured in the laboratory. They are only known through certain enzymes involved in sulfide oxidation and autotrophic carbon dioxide fixation, or their rRNA (5S, 16S) sequences demonstrating that they belonged to the gamma subdivision of Proteobacteria.

In the western Pacific, gastropods are members of vent communities. Although less studied, these snails also harbor autotrophic bacteria within their gill filaments.

An important question for these symbionts is the problem of transmission along their life cycles. Molecular techniques based on amplification (PCR) of



Fig. 12.4. A mussel (*Bathymodiolus*) bed at the Mid-Atlantic Ridge (courtesy of Daniel Desbruyères, IFREMER, Centre de Brest, France)

genes encoding for 16S rRNA, and in situ hybridization showed that a direct transmission through oocytes did exist for the bivalve *Calymene magnifica* (Cary and Giovannoni 1993). However, in the case of the tube worm *Riftia*, symbiont installation seems to occur after metamorphosis. This implies that the symbiont is free in the environment during at least one part of its life cycle. However, it has not yet been detected in environmental samples.

12.3.3 Polychaetous Annelids

Bivalve and tube worms live in areas where vent fluids, due to mixing with cold water, have moderate temperatures (10–20°C). But other invertebrates live close to hydrothermal chimneys and consequently are exposed to high temperatures. One individual of the polychaetous worm *Alvinella pompejana* (Chevaldonné et al. 1992) (Fig. 12.5) has even been pictured while twisted around a temperature probe at 100°C! The story did not say if the worm was happy or not! These animals live within the tubes they build onto the outer walls of active hydrothermal chimneys in the Eastern Pacific. Temperature of their environment is not permanently around 100°C, but more probably in the range 40–60°C. Nevertheless, *Alvinella* is probably one of the most thermotolerant metazoan. The animals harbor a very dense epibiotic microbial community on their outer



Fig. 12.5. Polychaetes *Alvinella pompejana* in their tubes at the East Pacific Rise (courtesy of Daniel Desbruyères, IFREMER, Centre de Brest, France)

integuments, and some epibionts are inserted at very precise locations, as confirmed by structural modification of the worm's cuticle. SEM and TEM observations showed various morphotypes for these epibionts (filaments, rods, cocci, twisted and stalked forms), and some of them have been successfully cultured in the laboratory (Prieur et al. 1989), showing their metabolic diversity (aerobes, anaerobes, nitrifiers, sulfate reducers, etc.).

Several heterotrophic bacterial strains have been isolated and appeared resistant (or multiresistant) to elevated concentrations of heavy metals such as cadmium, zinc, copper or silver. About 25% of the strains studied harbored one or several plasmids. Occurrence of plasmids with some similarities for strains belonging to different phyla could suggest the existence of genetic transfers in the vent environments. However, the dominant morphotypes that consist in filaments, visible with the naked eye, also escaped laboratory cultures. Analysis of their 16S rRNA revealed that they belong to the epsilon subdivision of Proteobacteria, a phylogenetic lineage, whose cultured known members are microaerophilic or anaerobic and metabolize sulfur compounds, or utilize hydrogen to reduce nitrate (Cary et al. 1997).

12.3.4 Crustaceans

Vent communities of the East Pacific Rise are characterized by tube worms and alvinellid worms, but Atlantic vents are unique since they show spectacular clouds of shrimp, swarming around active hydrothermal chimneys. These shrimp belong to several genera. The genus *Rimicaris* is one of the most studied. Mouth parts of these shrimps are enlarged and harbor epibiotic filament and rod-shaped bacteria. Using molecular techniques, it was shown that these symbionts also belong to the epsilon subdivision of Proteobacteria, but might also be found on sulfide rocks, even with no shrimp around (Polz and Cavanaugh 1985). For both annelid and crustacean symbiosis, the trophic role of the epibiotic bacteria has not been clearly demonstrated. More recently, SEM investigations of the digestive tract of *Rimicaris* shrimp revealed that rod-shaped bacteria were present in their hind gut, but their possible role in the digestive processes is still to be demonstrated.

The large invertebrate species, which depend on bacterial chemosynthetic symbiosis, are accompanied by smaller species and secondary consumers such as crabs or fish. Also, microbial communities not involved in symbiosis are abundant and diversified, although much less studied, with the exception of those forming bacterial mats or fouling various surfaces. But most microbiologists concentrated on the study of black smokers, hot fluids and associated microorganisms.

12.4 Life at High Temperature

Among physical parameters driving life, temperature is probably the most studied. First, heat was (and still is) the easiest mean by which to sterilize an object

or a liquid solution (120°C in an autoclave, 150°C in a dry oven). Consequently, it is not surprising that the first organism living optimally above 70°C was only discovered in the late 1960s. *Thermus aquaticus* was first isolated by Brock from a hot spring at Yellowstone National Park (USA). More than 30 years later, the enzyme DNA polymerase from *Thermus aquaticus* made possible DNA amplification through the polymerase chain reaction (PCR) with applications in basic biology, medical sciences, food microbiology and even criminology. In 1977, Brock isolated *Sulfolobus acidocaldarius*, the first hyperthermophilic organism (optimal temperature for growth above 80°C) (Brock 1985), a member of the third domain of life discovered the same year by Woese: the *Archaea*.

12.4.1 Novel Microorganisms in the Bacteria Domain

The occurrence of hot fluids venting at temperatures up to 350°C certainly lead microbiologists to focus on hyperthermophilic organisms. However, enrichment cultures of hydrothermal chimney samples at temperatures below 90°C allowed to isolate several organisms from the *Bacteria* domain (Table 12.2) (see Alain 2003 and Prieur 2004 for review). The first were strict anaerobes fermenting organic matter from the genus *Thermotoga* and *Thermosipho*. Still among anaerobes, an autotrophic organism, *Desulfurobacterium thermolithotrophum*, which utilized hydrogen as electron donor and thiosulfate as electron acceptor was described. Later, aerobic heterotrophic thermophiles were also described for the genera *Thermus* and *Bacillus*. These organisms were not expected since the thermal gradient is very steep: a few decimeters may separate the hot anaerobic zone from the cold oxygenated deep water, and the border line between anaerobiosis and aerobiosis is at about 30°C. However, certain hydrothermal chimneys have a porosity that allows cold water intake and results in an internal circulation of slightly oxygenated warm water at temperatures around 70°C. Further, relative genera with a chemoorganotrophic metabolism such as *Oceanothermus*, *Vulcanithermus*, *Caldithrix*, *Caloranaerobacter* were isolated and described. Thermophilic sulfate-reducers such as *Thermodesulfator indiensis*, or iron-reducers such as *Deferribacter abyssi* were also recently described.

To improve the knowledge of diversity of vent prokaryotes, molecular approaches (PCR amplification of 16S rRNA genes) were also used. For the *Bacteria* domain, this approach revealed the importance of the epsilon subgroup of *Proteobacteria*. Furthermore, culture methods allowed to isolate novel members from this lineage, belonging to the genera *Nautilia* and *Caminiibacter*. These organisms are anaerobes or microaerophilic, and may use nitrate or sulfur as electron acceptors

12.4.2 Novel Microorganisms in the Archaea Domain

Three lineages are distinguished for *Archaea*: *Euryarchaeota*, *Crenarchaeota* and the *Korarchaeota*. The latter lineage is only known from environmental 16SrRNA

Table 12.2. Novel Bacteria from deep-sea hydrothermal vents (A: autotroph; H: heterotroph; F: fermentative; S°: elemental sulfur required; in the energy column, compounds on the left of “/” are electron donors and compounds on the right of “/” are electron acceptors; μO_2 : microaerophilic; OC: organic carbon)

Lineage	Genus	Species	Carbon source	Energy	Optimal T (°C)
Aquificales	<i>Persephonella</i>	<i>marina</i>	A	H_2/NO_3^- , S° μO_2 , $\text{S}_2\text{O}_3^{2-}$	70
		<i>guaymasensis</i>	A	H_2/NO_3^- , S° μO_2 , $\text{S}_2\text{O}_3^{2-}$	80
		<i>hydrogenophila</i>	A	H_2/NO_3^- , μO_2 , $\text{S}_2\text{O}_3^{2-}$	70
Desulfurobacteriales	<i>Desulfobacterium</i>	<i>thermolithotrophum</i> <i>crinifer</i>	A	$\text{H}_2/\text{S}^\circ$ H_2/NO_3^- , S°	70 60–65
Thermotogales	<i>Marinitoga</i>	<i>camini</i>	H	F/S°	55
		<i>piezophila</i>	H	F, S°	65
	<i>Thermotoga</i>	<i>sp.</i>	H	F	80
	<i>Thermosipho</i>	<i>melanensis</i>	H	F/S°	70
		<i>japonicus</i>	H	F/S°	72
Thermodesulfurobacteriales	<i>Thermodesulfobacterium</i>	<i>hydrogenophilum</i>	A	$\text{H}_2/\text{S}^\circ$	75
<i>Thermus</i> /	<i>Thermus</i>	<i>sp.</i>	H	OC/O ₂	70–80
<i>Deinococcus</i>		<i>Thermophilus</i> GY1211	H	OC/O ₂	75

Table 12.2. (continued)

Lineage	Genus	Species	Carbon source	Energy	Optimal T (°C)
	<i>Marinithermus</i>	<i>hydrothermalis</i>	H	OC/O ₂	67
	<i>Vulcanithermus</i>				
	<i>Oceanithermus</i>	<i>profundus</i>	H/A	H ₂ , OC/O ₂	60
				μO_2	
Deferritiales	<i>Deferribacter</i>	<i>desulfuricans</i>	H	OC/S ^o , NO ³⁻ , As	60–65
		<i>abyssi</i>			
	<i>Bacillus</i>	<i>sp.</i>	H	F.; OC/O ₂	60–80
Firmicutes	<i>Caloranaerobacter</i>	<i>azorensis</i>	H	F	65
	<i>Camimicella</i>	<i>sporogenes</i>	H	F	55–60
Proteobacteria	<i>Camimibacter</i>				
	<i>Nautilia</i>				
	<i>Sulfospirillum</i>				
	<i>Thermodesulfatator</i>				
Novel lineage	<i>Caldithrix</i>	<i>abyssi</i>	H/A	Acetate, H ₂ /NO ₃ ⁻ , F	

sequences, and no cultured organism has been reported. Originally discovered in hot springs in Yellowstone National Park, they recently have been detected in some deep-sea hydrothermal vents. The *Euryarchaeota* gather organisms from various metabolic types (Table 12.3) (see Alain 2003 and Prieur 2004 for review). The most frequently isolated from deep-sea hydrothermal vents are the *Thermococcales*, with the genera *Pyrococcus* and *Thermococcus*, for a total of about 15 abyssal species. All are hyperthermophilic, grow optimally above 80°C, and ferment organic compounds and in particular, peptides. Most of them require elemental sulfur or a sulfur amino acid in the culture medium, and produce hydrogen sulfide. The dominance of *Thermococcales* in collections of hydrothermal microorganisms may be surprising, and does not necessarily correspond to the natural situation. Although the number of species described certainly correspond to a real diversity, it must be noticed that these organisms are relatively easy to grow in the laboratory. These organisms are strict anaerobes, but may resist oxygen exposure if it occurs at low temperature (2–4°C), or under organic carbon starvation. All these observations could explain why they apparently dominate in enrichment cultures and consequently in laboratory strain collections. Recently, a third *Thermococcales* genus, *Paleococcus ferrophilus* was described (Takai et al. 2000). This organism requires ferrous iron as electron acceptor in absence of elemental sulfur.

Still among *Euryarchaeota*, two sulfate-reducing species were described for the genus *Archaeoglobus* (Burggraf et al. 1990) as well as several methanogens (producing methane from hydrogen and carbon dioxide) (Jones et al. 1983). Close to *Archaeoglobus*, the genus *Geoglobus* can utilize ferric iron as an electron acceptor. The methanogens belong to the genera *Methanococcus*, renamed *Methanocaldococcus*, and *Methanopyrus*. For several years, *Methanopyrus kandleri*, also isolated from a shallow water hot spring in Iceland, remained the most thermophilic organism on Earth, growing optimally at 106°C, with a maximum growth temperature of 110°C (Kurr et al. 1991).

Crenarchaeota constitutes the second main lineage for *Archaea*. All are hyperthermophilic organisms metabolizing sulfur or sulfur-containing compounds. Several species have been isolated from deep-sea hydrothermal vents. They are strictly anaerobic, but show various metabolic types. They belong to the genera *Desulfurococcus*, *Staphylothermus*, *Pyrodictium* and *Pyrolobus*. The last genus includes the species *Pyrolobus fumarii*, which grows optimally at 110°C, with a maximum at 113°C (Blöchl et al. 1997). This is the most thermophilic organism presently known on Earth. Moreover, it can survive a two hours exposure in an autoclave (120°C). Very recently, a novel phylogenetic lineage has been proposed within the *Archaea*, with the discovery of *Nanoarchaeum equitans* (Huber et al. 2002), a very small-sized microorganism (400nm) living in very close association with cells of the Archaeon *Ignicoccus pacificus*.

Comparison of thermophilic and hyperthermophilic genera isolated from deep-sea hydrothermal vents and other hot environments shows rather few differences. However, acidophilic hyperthermophiles such as *Sulfolobus* have not been

Table 12.3. Novel Archaea from deep-sea hydrothermal vents (A: autotroph; H: heterotroph; F: fermentative; S°: elemental sulfur required; in the energy column, compounds on the left of “/” are electron donors and compounds on the right of “/” are electron acceptors; μO_2 : microaerophilic)

Lineage	Genus	Species	Carbon source	Energy	Optimal T (°C)
<i>Euryarchaeota</i>					
Methanococcales	<i>Methanocaldococcus</i>	<i>jannaschii</i>	A	H ₂ /CO ₂	85
		<i>infernus</i>	A	H ₂ /CO ₂	85
		<i>vulcanius</i>	A	H ₂ /CO ₂	80
	<i>Methanothermococcus</i>	<i>okinawensis</i>	A	H ₂ /CO ₂	60–65
Methanopyrales	<i>Methanopyrus</i>	<i>kandleri</i>	A	H ₂ /CO ₂	98
Thermococcales	<i>Thermococcus</i>	<i>guaymensis</i>	H	F, S°	88
		<i>aggregans</i>	H	F, S°	88
		<i>barossi</i>	H	F, S°	82
		<i>fumicolans</i>	H	F, S°	85
		<i>hydrothermalis</i>	H	F, S°	85
		<i>peptonophilus</i>	H	F, S°	85–90
		<i>siculi</i>	H	F, S°	85
		<i>chitonophagus</i>	H	F, S°	85
		<i>barophilus</i>	H	F, S°	85
		<i>atlanticus</i>	H	F, S°	85
		<i>gammatolerans</i>	H	F, S°	88
	<i>Pyrococcus</i>	<i>abyssi</i>	H	F, S°	96
		<i>glycovorans</i>	H	F, S°	95
		<i>horikoshii</i>	H	F, S°	98

Table 12.3. (continued)

Lineage	Genus	Species	Carbon source	Energy	Optimal T (°C)
	<i>Paleococcus</i>	<i>ferrophilus</i>	H	F, S° Fe ⁺⁺⁺	83
Archaeoglobales	<i>Archaeoglobus</i>	<i>profundus</i>	H,A	Required H ₂ , OC/SO ₄ ²⁻	82
		<i>veneficus</i>	H,A	H ₂ , OC/SO ₃ ²⁻ S ₂ O ₃ ²⁻	75–80
	<i>Geoglobus</i>	<i>ahangari</i>	H,A	H ₂ , OC/Fe ³⁺	88
Crenarchaeota					
Desulfurococcales	<i>Ignicoccus</i>	<i>pacificus</i>	A	H ₂ /S°	90
	<i>Staphylothermus</i>	<i>marinus</i>	H	F/S°	85–92
	<i>Pyrodicticum</i>	<i>abyssi</i>	H	F/S°	97
	<i>Pyrolobus</i>	<i>fumarii</i>	A	H ₂ /NO ₃ ³⁻ μO ₂ , S ₂ O ₃ ²⁻	106
“Nanoarchaeota.”	<i>Nanoarchaeum</i>	<i>equitans</i>	A?	?	90

reported from deep-sea vents. Although hydrothermal fluids are acidic, they are quickly diluted into seawater (pH 8). Consequently a hot (less than 113°C) and acidic (pH < 5 to 6) biotope has not been identified yet. Some species such as *Staphylothermus marinus* have been reported for shallow and deep hot springs, and probably have an ubiquitous distribution. For most of these hyperthermophiles, novelty was often found at the species level as confirmed by the numerous species called “*profundus*”, “*abyssi*” or “*hydrothermalis*”. However, several genera seem to be present only in deep-sea vents (*Marinitoga*, *Pyrolobus*, *Paleococcus* or *Desulfurobacterium*). But it is not firmly demonstrated that they only thrive in deep environments, and perhaps shallow vent species are still waiting to be discovered.

The existence of very hot fluids certainly drove the search for hyperthermophiles in deep-sea hydrothermal vents. But, if the maximum temperature limit for life (113°C) has been reported for a deep vent organism (*Pyrolobus fumarii*), it only overcomes by three degrees Centigrade the maximum growth temperature of *Methanopyrus kandleri*, isolated also from a 100m-deep vent in Iceland, or *Pyrodicticum occultum* isolated from a shallow vent at Vulcano Island (Sicilia, Italy).

12.5 Response to Hydrostatic Pressure

Among all physical parameters mentioned above, hydrostatic pressure is characteristic of all deep vent sites. Responses to elevated hydrostatic pressure have been studied for microorganisms from cold deep environments. Barophilic and even obligate barophilic organisms have been described, and their physiology and adaptations studied at the molecular level (Yayanos 1986; Prieur and Marteinsson 1998; Kato 1999). But much less is known for deep-sea vent thermophiles. One reason is that the collection of a black smoker sample, using a retaining-pressure collecting device is not easy: a single, very small metallic particle (frequent in the vent plume) may cause a leak in the sampler. For these reasons, all the samples studied have been decompressed when brought up to the surface. All enrichment cultures that lead to the isolation of organisms described above have been carried out under atmospheric pressure, or a slight hyperbaric pressure to avoid boiling of culture media for temperatures close or above 100°C.

However, the question of the effect of elevated hydrostatic pressure on deep-sea thermophiles is still open. Several microbiologists exposed deep-sea thermophilic organisms isolated under atmospheric pressure to elevated hydrostatic pressure, using pressurized bioreactors. Several responses were observed (Prieur and Marteinsson 1998): some organisms were barosensitive, and showed a slower growth rate under pressure; some others were barotolerant, and their growth was not affected by pressure. But most of the species studied appeared to be barophiles, and their growth rates were enhanced by hydrostatic pressure. Moreover, their optimal growth temperatures increased by a few degrees Centigrade.

For instance, the maximum growth temperature for *Pyrococcus abyssi* is 102°C, but it reached 105–106°C under elevated hydrostatic pressure (Erauso et al. 1993). Similarly, optimum growth temperature increased from 96 to 100°C. Similar observations, reported by several authors were compiled and discussed by Deming and Baross (1993) who noted that for all barohyperthermophiles, pressure allowing optimal growth was always above the pressure existing at capture depth (for an organism isolated from a sample collected at 2000m depths, optimum pressure for growth was 40MPa). This point is very remarkable since for baropsychrophiles, optimum pressure for growth was always under the pressure existing at capture depth (Yayanos 1986).

When they reported that barohyperthermophiles had an optimum growth pressure above in situ pressure, Deming and Baross (1993) asked the question: are black smokers windows to a deep biosphere? This concept of a deep biosphere has been progressively confirmed by the discovery of viable bacteria in deep samples from various origins: deep aquifers, deep rocks, continental and offshore oil reservoirs, deep marine sediments (several hundred meters), etc.

Another experimental approach consisted in isolating hyperthermophilic organisms from enrichment cultures (and subcultures) carried out under elevated hydrostatic pressure. This method isolated a novel *Archaea* species (*Thermococcus barophilus*) and a novel Bacteria species (*Marinitoga piezzophila*) (Marteinsson et al. 1999; Alain et al. 2002). For both organisms, which are not real taxonomic novelties, but only novel species, growth rates were enhanced under hydrostatic pressure, but growth remained possible under atmospheric pressure. In the case of *Marinitoga*, cell divisions were very slow under atmospheric pressure and cell morphology was clearly modified. Maximum growth temperature was again increased, but optimal temperature was not affected by pressure. Further information was obtained from a study of total proteins of *T. barophilus* cultured under elevated hydrostatic pressure or atmospheric pressure. Under hydrostatic pressure, an unknown 35kDa protein was expressed. Under atmospheric pressure, a 60kDa protein was expressed, which corresponded to a stress protein already known for other hyperthermophilic *Archaea*. All these data confirm that this *Thermococcus* species is a true barothermophile (Marteinsson et al. 1999).

12.6 Other Specific Adaptations

12.6.1 Fluctuations of Environmental Conditions

At deep-sea hydrothermal vents, several parameters may fluctuate and affect the physiology of microorganisms. This is the case for oxygen and nutrient concentrations. Hyperthermophiles are mostly anaerobic and sensitive to oxygen exposure. This situation may occur in situ, as a result of a decreasing venting activity. Using *P. abyssi* as a model strain, it was shown that if oxygen was toxic for this strain at growth temperature, *P. abyssi* was able to survive for several weeks at 4°C in an oxygenated environment (Marteinsson et al. 1997). Similarly,

this organism was not affected by starvation in a minimal medium (no organic carbon) for at least one month at 4°C, and only minimally affected at 95°C for several days. Furthermore, cells were more resistant to oxygen under starvation conditions. These results may indicate that, at least *Pyrococcus* but probably other vent organisms are adapted to fluctuating conditions taking place in vent environments, and that as a consequence they might be able to spread from one vent field to another.

12.6.2 Heavy Metals

Several heterotrophic strains (*Acinetobacter*, *Alteromonas*, *Pseudomonas*, *Vibrio*) have been isolated and appeared resistant (or multiresistant) to heavy metals such as cadmium, zinc, silver, arsenic and particularly copper. About 20% of the strains studied harbored one or several plasmids (up to five) of sizes from 4.6 to 157kb. Occurrence of plasmids with some similarities for strains belonging to different phyla could suggest existence of genetic transfers in the vent environments (Jeanthon 2000).

12.6.3 Ionizing Radiations

Several hyperthermophilic Archaea have been shown to resist ionizing radiations but not in the same range as *Deinococcus radiodurans*, the most radioresistant microorganism known so far (10 to 20kGy while eukaryotic cells are killed at 5Gy (Battista 1997)). In the vent environment, it was reported that invertebrates living on black smokers were exposed to natural radiations one hundred times above those received by man on Earth. For these reasons, responses of the Archaeon *Pyrococcus abyssi* to ionizing radiations have been studied (Gerard et al. 2001). It was observed that this organism did not show any loss of viability until 2kGy of g-irradiation. It was then established that *P. abyssi* did not have a specific DNA repair system that could explain its radioresistance although the chromosomal DNA was fully restored under optimal growth conditions.

To isolate more radioresistant hyperthermophiles, enrichment cultures under conditions designed for Thermococcales growth were exposed to elevated doses of radiations (20 to 30kGy), and then processed for strain isolation. Three radioresistant strains were isolated and fully described as novel species: *Thermococcus gammatolerans* (about the same range of resistance as *D. radiodurans*), *Thermococcus radiotolerans* and *Thermococcus marinus* (Jolivet et al. 2003).

12.7 Lessons from Microbiology of Hydrothermal Vents

If the study of deep-sea hydrothermal vents remains a domain of scientific research for a rather small number of scientists (these remote environments are

difficult to explore and sample), outreach initiatives about these environments have been very active and many people have watched on TV the black smokers, the deep-sea submersibles, or even read about hydrothermal vents in police fiction or cartoon books.

However, even if the deep-sea vents are now part of the human knowledge, it must be recalled that they were discovered in 1977, less than 30 years ago.

This is the **first lesson** to keep in mind: the biological exploration of planet Earth is not completed yet. If it is dramatically true that the known biodiversity decreases, a consequence of the modification of many biotopes by human activities, it is also true that a tremendous number of species is still to be discovered, particularly in the microbial world. It was reported that about 5000 to 6000 prokaryotic species were described, but probably 100 000 to 200 000 remain to be discovered (Madigan et al. 2003).

The **second lesson** concerns the physiochemical limits for life. Thirty years ago, the highest temperature for life was only about 70°C. Since that time, microbiologists have considerably extended our knowledge of the limits of life.

At the bottom of the temperature scale, it is difficult to make the difference between life (briefly all processes concluded by cell division) and survival (no cell division, but potentiality for future cell division preserved). For instance prokaryotic, but also eukaryotic cells are routinely preserved (no growth) in liquid nitrogen (−196°C). Glacier bacteria (psychrophiles) are thought to divide once a year. But this is an average and it is not for all cases documented if the division process takes a long time, or a short time; occurring once a year when the environmental conditions are the most favorable.

At the top of the temperature range, 106°C is the highest optimal temperature for growth reported for a living organism (*Pyrolobus fumarii*), and this organism continues to grow at 113°C (Blöchl et al. 1997). This value represents the highest temperature allowing life, admitted by the scientific community. However, it has been recently reported that microorganisms could grow up to 121°C. But this result was criticized because the methods used for cell counting (and consequently determining growth, and growth rate) were not accurate enough: it was probably a case of survival at very high temperature, similarly to *Pyrolobus fumarii*, which survives one hour in an autoclave at 120°C. Finally, the organisms studied were not isolated in pure culture, no novel species was described, and consequently deposited in a type culture collection (Kashefi and Lovley 2003). The question is still open.

pH

Living microorganisms have been isolated and described from alkaline (pH 10–11) or acidic environments (pH 1 or < 1) (Madigan et al. 2003). If acidophiles were intensively studied, alkaliphiles are less known, and probably novel organisms should be discovered, particularly if pH and elevated temperature are considered together.

Salinity

The effect of salt concentration is rather well-documented, and extreme halophiles (*Archaea* domain) are well-known and studied (Madigan et al. 2003). However, novelties could arise if, again, salt concentration was considered together with another parameter such as elevated temperature and hydrostatic pressure.

Pressure

Microbial exploration of deep-sea hydrothermal vents has particularly contributed to thermophily with the discovery of the most hyperthermophilic organism (*Pyrolobus fumarii*, maximum temperature for growth: 113°C). But also organisms that combine adaptations to several extreme conditions have been discovered. For instance, all the deep-sea vent thermophiles and hyperthermophiles are at least piezotolerant, since they are exposed to hydrostatic pressure in their natural environments. Some of them are even happier when hydrostatic pressure increases (*Thermococcus barophilus*, *Marinitoga piezophila*). *Thermococcus gammatolerans* grows optimally at 85°C, normally lives at a depth of 2000m (20MPa), and resists ionizing radiations of about 10kGy. It has even been suggested that this organism could resist high concentrations of heavy metals and experiments about this point are currently in progress. Finally, it must be pointed out that, despite the numerous and exciting discoveries of these last years, the physicochemical limits for life are not yet known.

When examining the physicochemical limits for life, we only considered conditions allowing a given organism to carry out its complete life cycle, or to cope with conditions lethal for other organisms (for instance radiation resistant microorganisms have highly efficient DNA repair mechanisms). Beyond these limits, organisms are able to survive in a “dormant” stage. The best known dormant stage is the spore of certain gram-positive Bacteria; it is also the most efficient mechanism discovered so far. Bacterial spores have been found in very old samples, and a viable spore of a bacterium of 250 million years old has even been reported (Vreeland et al. 2000). Dormant stages different from the spores, probably also exist for other Bacteria and Archaea lineages, but have yet to be reported.

The **third lesson** concerns energetic metabolisms, but probably requires that some definitions must be reminded before being expressed.

One of the cell characteristics is reproduction, or growth, which implies that a cell may build all its components from elements taken up from its environment. Transformation into cell components of elements taken up from the environment is called anabolism or biosynthesis. This biosynthesis requires energy, also required for other functions such as nutrient transport, waste elimination or cell motility. Consequently, each cell must be able to gain energy from its environment, either from light or chemical reactions. These chemical reactions mostly consist in degradation of complex molecules in more simple molecules, this degradation giving energy. These processes are named catabolism or energetic metabolism. All these chemical processes present together in cells (an-

abolism plus catabolism) constitute the metabolism (Madigan et al. 2003; Prieur 2003).

Living organisms are classified in several groups, depending on their energetic metabolism. Those using light as an energy source are called phototrophs. Those gaining their energy from chemical reactions are named chemotrophs. In this case, those using organic molecules as energy sources are chemoorganotrophs, while those using inorganic molecules as energy sources are called chemolithotrophs. During these chemical reactions, electron(s) is (are) transferred from the energy source (electron donor) to an electron acceptor. If this electron acceptor is molecular oxygen, the organism using this process (whatever the electron donor is) is an aerobe. If the electron acceptor is any molecule (sulfate, nitrate, ferric iron, etc.), but not molecular oxygen, the organism using this process is an anaerobe. Also, in some cases, the electron acceptor is provided internally during the degradation of the organic substrate through metabolic processes called fermentations. Some organisms can use oxygen as an electron acceptor or an alternative electron acceptor, or carry out a fermentation: they are named facultative anaerobes. It must be recalled that molecular oxygen present in the Earth's atmosphere, or dissolved in water masses (lakes, rivers, oceans) is a product of oxygenic photosynthesis. In this case, light energy is used to photolyse water, which produces a reducing power to reduce carbon dioxide into organic carbon (sugar), and consequently releasing oxygen. This photosynthetic process is carried out by plants, seaweeds, microalgae and certain *Bacteria*: the *Cyanobacteria*. Some other *Bacteria* also carry out photosynthesis, but in a different way: reducing power required for carbon dioxide reduction comes from the environment (hydrogen sulfide, etc.), and no oxygen is released during this process called anoxygenic photosynthesis.

The study of microorganisms living in deep-sea hydrothermal vents has pointed out that chemotrophy, and essentially chemolithotrophy was a very efficient metabolic process (lithoautotrophic symbionts support abundant communities of invertebrates and many free-living lithotrophs have doubling times of less than 30 minutes), and that a tremendous variety of combinations of inorganic electron donors and electron acceptors (other than dioxygen for the latter) could sustain life, even in severe environmental conditions such as elevated temperature and pressure (Prieur 2004). All combinations have not been tested yet; in particular the diversity of metals to be used as electron acceptors should be further investigated. It must also be pointed out that some metabolisms, not reported yet for deep-sea vent (thermophilic or not) Prokaryotes, have been recently demonstrated such as anaerobic methane oxidation or anaerobic ammonium oxidation (anammox) in deep-sea sediments and sewage treatment plants, respectively. If the latter was proved to exist at high temperatures, it would allow the nitrogen cycle to *work* at high temperatures, since aerobic ammonium oxidizers have not been detected above 60°C.

Many Prokaryotes, from a variety of environments, can use several chemical reactions to gain their energy. In other words, versatility is the rule: according

to environmental conditions, a given prokaryotic cell may switch from aerobic respiration to anaerobic respiration or fermentation, from lithoautotrophy to mixotrophy (energy gained from inorganics but organic carbon sources), even to photoautotrophy or heterotrophy. If this wide and versatile metabolic diversity is coupled to the extraordinary physiological capacities of many organisms to colonize extreme (and not extreme) environments, it becomes evident that there is an unlimited amount of biotopes on Earth, suitable for prokaryotic life: we have again strong arguments to support the first lesson statement that biological exploration of planet Earth is far from complete.

The fourth lesson seems to be obvious, but sometimes, even microbiologists forget the point: microorganisms are small, generally about one micrometer. Then, a microbial biotope must be considered at the microbial scale. Imagine a man or a complex vertebrate (mammal, bird, etc.) moving in a given direction for 3km. How many environments, landscapes, biotopes does he encounter during such a journey (grass, some trees, meadows, stream, dry path, gravels, etc.)? For a prokaryotic cell, the same journey will take place within two millimeters, and it will meet anaerobic, organic rich, humid, dry, aerobic conditions (Madigan et al. 2003). For instance, several years ago, from the same small amount (a few grams) of hydrothermal chimney material, we obtained after different enrichment cultures, a mesophilic chemolithoautotrophic aerobic bacterium (*Thiobacillus hydrothermalis*), and an hyperthermophilic heterotrophic anaerobic archaeon (*Pyrococcus abyssi*) (Erauso et al. 1993).

The final lesson concerns exploration of the solar system, and beyond, for detection of past or present life.

Terrestrial Prokaryotes must serve as references for the search of life in the solar system, and particularly Mars and Europa. They must serve as references for their physiological adaptations (temperature, salt, pH, etc.), their survival processes, their biotope location (particularly subsurface environments), and their energetic metabolisms (energy sources, combination of electron donors and acceptors).

If life does not presently occur on Mars, but had occurred on early Mars when life arose on Earth, the probable metabolic history of life on Earth should be taken as a model: anaerobic respiration of inorganic compounds (anaerobic lithotrophy) preceded anoxygenic and latter oxygenic photosynthesis. For these primitive lithotrophic microorganisms, carbon dioxide fixation is not always carried out and uptake of already synthesized organic molecules (mixotrophy) is common. Such metabolisms, although known for a long time, were re-discovered with the exploration of hydrothermal vents and their microbial communities. Hydrothermal environments should be considered as focused targets, since these environments provide energy, carbon (event organics) sources and electron acceptors, and a variety of inorganic surfaces suitable for biofilm formations (and molecule adsorption), and a thermal gradient that allows processes requiring different temperature conditions to co-exist within a short distance (centimeters to decimeters).

It is more difficult to adapt these lessons to Europa, since the exploration of this moon of Jupiter, and particularly the subice environment is far from being scheduled. If an ocean of liquid water does exist on Europa, only one of the requirements for life is present (water). Carbon (organic?) and energy sources are still to be identified. However, terrestrial Prokaryotes will again serve as references, and studies under progress on Vostok Lake ice and similar biotopes, will certainly serve as models for Europa investigations (Petit 2003).

To increase the possibility to detect such life, techniques and instruments will be designed to measure very small amounts of biosignatures. But do we know all possible biosignatures? Certainly not. We now know a variety of organisms that must serve as references, but we do not know all their biosignatures. A catalogue of reference biosignatures should be established, and this will require a strong cooperation between microbiologists (they know microbial diversity), geologists (they know about rock formation, mineralogy, and fossilization), and organic chemists (they know how to detect small amount of specific molecules).

The search for past or present life beyond the solar system, must integrate our knowledge about physiology and metabolism of actual terrestrial Prokaryotes. Particularly, the lack of molecular oxygen in the atmosphere of a planet does not mean that life does not exist. It only means, if other physicochemical conditions are acceptable, that aerobic respiration and probably oxygenic photosynthesis do not exist. In other words, if a remote civilization had observed and analyzed Earths atmosphere three billion year ago, these individual would have probably claimed that Earth was not favorable for life, because of the very low oxygen concentration. An intensive study of isotopic fractionation of molecules released by metabolic processes (methane, nitrogen, nitric and nitrous oxides, hydrogen, sulfides, etc.) should be carried out in cooperation with geochemists and microbiologists to suggest reliable life signatures.

About 3.6Ga after life arose on Earth, its principal characteristic is diversity. We should not forget that point when looking around for another living planet in the universe. Extraterrestrial life might be very different from Earth organisms.

References

- Alain K (2003) Approches culturelles et moléculaires des assemblages microbiens associés aux Polychètes hydrothermaux de la famille Alvinellidae. Dissertation, l'université de Brest (France).
- Alain K, Marteinsson VT, Miroshnichenko ML, Bonch-Osmolovskaya EA, Prieur D, Birrien JL (2002) *Marinitoga piezophila* sp. nov., a rod-shaped, thermo-piezophilic bacterium isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* **52**: 1331–1339.
- Battista JR (1997) Against all odds: the survival strategies of deinococcus radiodurans. *Annu Rev Microbiol* **18**: 203–224.
- Blöchl E, Rachel R, Burggraf S, Hafenbradl D, Jannasch HW, Stetter KO (1997) *Pyrolobus fumarii*, gen. nov., sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113°C. *Extremophiles* **1**: 14–21.

- Brock TD (1985) Life at high temperatures. *Science* **230**: 545550.
- Burggraf S, Jannasch HW, Nicolaus B, Stetter KO (1990) *Archaeoglobus profundus*, sp. nov., represents a new species within the sulfate-reducing archaeobacteria. *Syst Appl Microbiol* **13**: 24–28.
- Cary SC, Giovannoni SJ (1993) Transovarial inheritance of endosymbiotic bacteria in clams inhabiting deep-sea hydrothermal vents and cold seeps. *Proc Natl Acad Sci USA* **90**: 56955699.
- Carry SC, Cottrell MT, Stein JL, Camacho F, Desbruyères D (1997) molecular identification and localization of filamentous symbiotic bacteria associated with the hydrothermal vent annelid *alvinella pompejana*. *Appl Environ Microbiol* **63**: 1124–1130.
- Cavanaugh CM, Gardiner SI, Jones ML, Jannasch HW, Watrebury JB (1981) Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: possible chemoautotrophic symbionts. *Science* **213**: 340–341.
- Cavanaugh CM (1983) Symbiotic chemoautotrophic bacteria in marine invertebrates from sulfide-rich habitats. *Nature* **302**: 58–61.
- Cavanaugh CM, Wirsén CO, Jannasch HW (1992) Evidence for methylotrophic symbionts in a hydrothermal vent mussel (*Bivalvia:Mytilidae*) from the Mid-Atlantic Ridge. *Appl Environ Microbiol* **58**: 3799–3803.
- Charlou JL, Donval JP, Fouquet Y, Jean-Baptiste P, Holm, N (2002) Geochemistry of high H₂ and CH₄ vent fluids issuing from ultramafic rocks at the rainbow hydrothermal field (36°14'N, Mar Mol Biol Biotechnol). *Chem Geol* **191**: 345–359.
- Chevaldonné P, Desbruyères D, Childress JJ (1992) Some like it hot and some even hotter. *Nature* **359**: 593–594.
- Corliss JB, Dymond J, Gordon LI, Edmond J, von Herzen RP, Ballard RD, Green K, Williams D, Bainbridge A, Crane K, van Andel TH (1979) Submarine thermal springs on the Galapagos Rift. *Science* **203**: 1073–1083.
- Deming JW, Baross JA (1993) Deep-sea smokers: windows to a subsurface biosphere. *Geochim Cosmochim Acta* **57**: 3219–3230.
- Durand P, Reysenbach A-L, Prieur D, Pace N (1993) Isolation and characterization of *Thiobacillus hydrothermalis* sp. nov., a mesophilic obligately chemolithotrophic bacterium isolated from a deep-sea hydrothermal vent in Fiji Basin. *Arch Microbiol* **159**: 39–44.
- Edmond J, van Damm K, MacDuff R, Measures C (1982) Chemistry of hot springs on the East Pacific Rise. *Nature* **297**: 187–191.
- Erauso G, Reysenbach AL, Godfroy A, Meunier JR, Crump B, Partensky F, Baross JA, Marteinsson V, Barbier G, Pace NR, Prieur D (1993) *Pyrococcus abyssi* sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Arch Microbiol* **160**: 338–349.
- Felbeck H (1985) CO₂ fixation in the hydrothermal vent tube worm *Riftia pachyptila* (Jones). *Physiol Zool* **58**: 272–281.
- Fiala G, Stetter KO, Jannasch HW, Langworthy TA, Madon J (1986) *Staphylothermus marinus* sp. nov. represents a novel genus of extremely thermophilic submarine heterotrophic archaeobacteria growing up to 98°C. *Syst Appl Microbiol* **8**: 106–113.
- Gérard E, Jolivet E, Prieur D and Forterre P (2001) DNA protection is not involved in the radioresistance of the hyperthermophilic archaea *Pyrococcus abyssi* and *Pyrococcus furiosus*. *Mol Genet Genom* **266**: 72–78.

- Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**: 63–67.
- Jannasch HW, Taylor CD (1984) Deep-sea microbiology. *Ann Rev Microbiol* **38**: 487–514.
- Jeanthon C (2000) Molecular ecology of hydrothermal vent microbial communities. *Antonie van Leeuwenhoek* **77**: 117–133.
- Jolivet E, Corre E, L'Haridon S, Forterre P, Prieur D (2003) *Thermococcus gamma-tolerans* sp. Nov., a hyperthermophilic archaeon from deep-sea hydrothermal vent that resists ionizing radiation. *Int J Syst Evol Microbiol* **53**: 847–851.
- Jones WJ, Leigh JA, Mayer F, Woese F, Woese CR, Wolfe RS (1983) *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. *Arch Microbiol* **136**: 254–261.
- Kashefi K, Lovley DR (2003) Extending the upper temperature limit for life. *Science* **301**: 934–935.
- Kato C (1999) Barophiles (piezophiles). In: Horikoshi K, Tsujii K (eds.) *Extremophiles in Deep-Sea Environments*, pp. 91–111. Springer, Berlin Heidelberg New York.
- Kurr M, Huber R, König H, Jannasch HW, Fricke H, Trincone A, Kristjansson JK, Stetter KO (1991) *Methanopyrus kandleri*, gen. and sp. nov. represents a novel group of hyperthermophilic methanogens, growing at 110 °C. *Arch Microbiol* **156**: 239–247.
- Le Pennec M, Prieur D (1984) Observations sur la nutrition d'un Mytilidae d'un site actif de la dorsale du Pacifique oriental. *CR Acad Sci Paris* **298**: 493–498.
- Madigan MT, Martinko JM, Parker K (2003). *Brock Biology of Microorganisms*. Pearson Education, Upper Saddle River, NJ, p. 1019.
- Marteinsson VT, Moulin P, Birrien JL, Gambacorta A, Vernet M, Prieur D (1997) Physiological responses to stress conditions and barophilic behavior of the hyperthermophilic vent Archaeon *Pyrococcus abyssi*. *Appl Environ Microbiol* **63**: 1230–1236.
- Marteinsson VT, Reysenbach AL, Birrien JL, Prieur D (1999) A stress protein is induced in the deep-sea barophilic hyperthermophile *Thermococcus barophilus* when grown under atmospheric pressure. *Extremophiles* **3**: 277–282.
- Marteinsson VT, Birrien JL, Reysenbach A-L, Vernet M, Marie D, Gambacorta A, Messner P, Sleytr U, Prieur D (1999) *Thermococcus barophilus* sp. Nov., a new barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *Int J Syst Bacteriol* **49**: 351–359.
- Petit JR, Blot M, Bulat S (2003) Le lac Vostok: à la découverte d'un environnement sous glaciaire et de son contenu biologique. In: Gargaud M (ed.) *Les traces du vivant*, pp. 275–318. Presses Universitaires de Bordeaux, Bordeaux.
- Polz MF, Cavanaugh CM (1995) Dominance of one bacterial phylotype at a Mid-Atlantic Ridge hydrothermal vent site. *Proc Natl Acad Sci USA* **92**: 7232–7236.
- Prieur D (1992) Physiology and biotechnological potential of deep-sea bacteria. In: Herbert RA, Sharp RJ (eds.) *Molecular Biology and Biotechnology of Extremophiles*, pp. 163–202. Blackie, New York.
- Prieur D, Marteinson VT (1998) Prokaryotes living under elevated hydrostatic pressure. *Adv Biochem Eng Biotechnol* **61**: 23–35.
- Prieur D (2003) Diversité des métabolismes. In: Gargaud M (ed.) *Les traces du vivant*, pp. 243–254. Presses Universitaires de Bordeaux, Bordeaux.

- Prieur D (2004) Microbiology of deep-sea hydrothermal vents: lessons for Mars exploration. In: Tokano T (ed.) *Water on Mars and Life*, pp. 299–318. Springer, Berlin Heidelberg New York.
- Prieur D, Benbouzid-Rollet N, Chamroux S, Durand P, Erauso G, Jacq E, Jeanthon C, Mevel G, Vincent P (1989) Distribution de divers types métaboliques bactériens sur un site hydrothermal profond (Dorsale de Pacifique Oriental à 13°N). *Cahiers Biol Mar* **30**: 515–530.
- Takai K, Sugai A, Itoh T, Horikoshi K (2000) *Palaeococcus ferrophilus* gen. nov., sp. nov., a barophilic, hyperthermophilic archaeon from a deep-sea hydrothermal vent chimney. *Int J Syst Evol Microbiol* **50**(2): 489–500.
- 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.
- Woese C (1987) Bacterial evolution. *Microbiol Rev* **51**: 221–271.
- Yayanos AA (1986) Evolution and ecological implications of the properties of deep-sea barophilic bacteria. *Proc Natl Acad Sci USA* **83**: 9542–9546.

13 Comets, Titan and Mars: Astrobiology and Space Projects

Yves Bénilan and Hervé Cottin

Introduction

All the solar system bodies were formed about 4.6 billion years ago, when a molecular cloud collapsed. The Sun, the telluric planets, the giant planets as well as their satellites, and the small bodies (asteroids and comets), were born from this material, which was probably quite homogeneously mixed in the beginning. From then, the physicochemical evolution yielded life at least on the Earth and also maybe on Mars or more speculatively on Europa. Thanks to space exploration, we try to figure out what makes the Earth such a peculiar object that it is the only one known to harbor life, and therefore, the extent to which life may have arisen somewhere else in the universe. To date, two factors appear to be of prime importance: a source of organic molecules and liquid water, the only solvent in which advanced molecular complexity appears to be possible to achieve, a preliminary condition for life. Henceforth, space missions dealing with astrobiology questions are conceived to answer these questions:

- What is the origin and the distribution of the organic matter in the solar system?
- Does water exist or did water exist in the liquid form on other objects?

The goal of this chapter consists of a brief review of space exploration dealing with three bodies of great astrobiological interest in our solar system: comets, Titan and Mars. Each of them is located at a specific level of our understanding of the origin of life on Earth, and the possibility that life could arise somewhere else. Comets as exogenous sources for prebiotic molecules, Titan as a model of an atmosphere in which endogenous syntheses of such molecules are very efficient, and finally Mars, which was the Earth's "twin-sister planet" more than three billions years ago, where it is now established that liquid water once flowed, and thus might have once harbored life.

13.1 An Astrobiological Look at the Solar System

13.1.1 The Origin of the Organic Matter

13.1.1.1 The Exogenous Track

Figure 13.1 is a scheme of the evolution of matter from our natal molecular cloud, to the different objects of the solar system. It shows, for each of them, the critical parameters that one considers favorable or unfavorable for life to

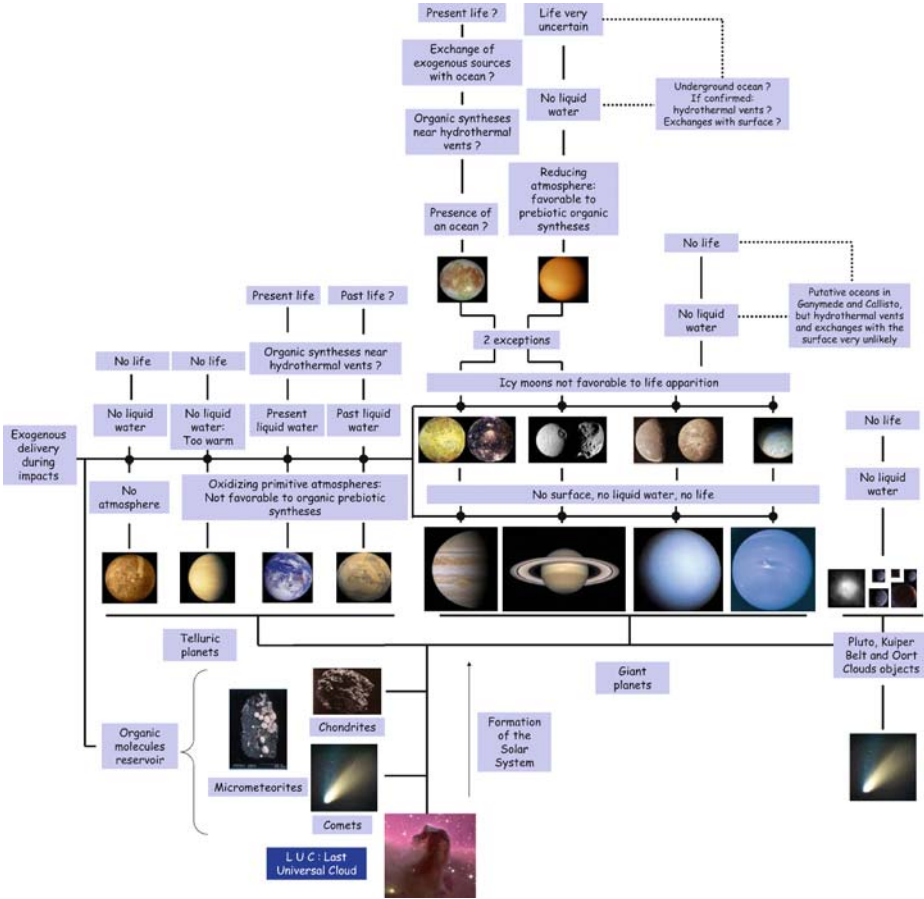
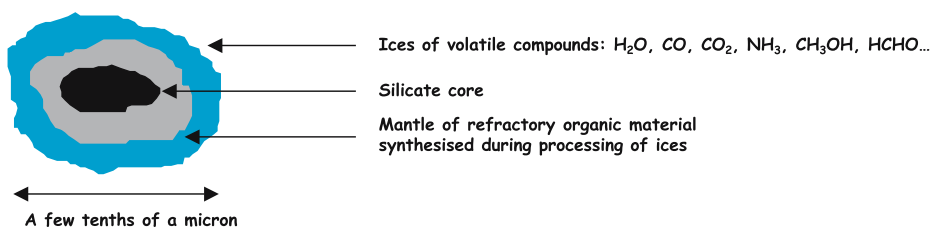


Fig. 13.1. Astrobiological tree of the solar system. The evolution of matter (see also Fig. 13.3) and astrobiological relevance of solar system objects. From *bottom to top*, and from *left to right*: the Horsehead nebula, comet C/1995 O1 (Hale-Bopp), Orgueil meteorite, Mercury, Venus, the Earth, Mars, Jupiter, Saturn, Uranus, Neptune, Pluto and objects of the Kuiper Belt and the Oort cloud, Io, Callisto, Mimas, Phoebe, Ariel, Oberon, Triton, Europa and Titan

appear on those bodies. Everything starts with a molecular cloud (Fig. 13.1), which one might call LUC (last universal cloud) as an analogy with LUCA (last universal common ancestor) the common ancestor to all living organisms on the Earth, and LUCY who once was the oldest found human ancestor (see Chap. 11). Most certainly, the molecular clouds that we are now able to observe in our galaxy show strong analogies with the composition of our natal cloud LUC. Remote sensing observations have shown that organic chemistry is very active and advanced in such clouds, as well in the gaseous phase as in the solid phase (interstellar ices) (Ehrenfreund and Charnley 2000). HCN, HC_3N or HCHO have been detected, and such molecules are of great astrobiological interest (Brack 2003; Oro and Cosmovici 1997).¹ Moreover, a large number of experimental simulations in the laboratory predict that one might expect the existence of molecules much more complex than the ones already detected. Such experiments consist in simulating the chemistry occurring in interstellar ices, when submitted to irradiations (with photons or charged particles) or thermal cycles (Despois and Cottin 2005). In molecular clouds, a large amount of organic matter should be frozen on condensation nuclei made of silicates (Fig. 13.2). Some gravitational perturbation probably leads the cloud to collapse on itself, which resulted in the birth of our Sun and our planetary system.

A schematic view consists of considering that the original composition of interstellar grains was lost since they have been either incorporated in the Sun or planets, or pyrolysed in the vicinity of the Sun. Interstellar ices could also sublimate in the warmest parts of the nebula in which they have been brought by turbulent radial mixing in the solar nebula (Fig. 13.3). However, the extension of the turbulent region determines the possibility to keep pristine interstellar matter in comets. Inside the turbulent region, radial mixing brings interstellar



Dust Particle in a Molecular Cloud

Fig. 13.2. Interstellar grain model. A silicate core on which volatile molecules condense in molecular clouds. The chemistry between those molecules leads to the synthesis of other compounds, with enhanced complexity. Our solar system formed with the accretion of such material. Some comets might have kept intact some of those grains

¹ In liquid water HCN chemistry leads to the synthesis of amino acids and puric bases (adenine and guanine), whereas HC_3N chemistry produces pyrimidic bases, and HCHO chemistry results in sugars or amino acids (when coupled with HCN).

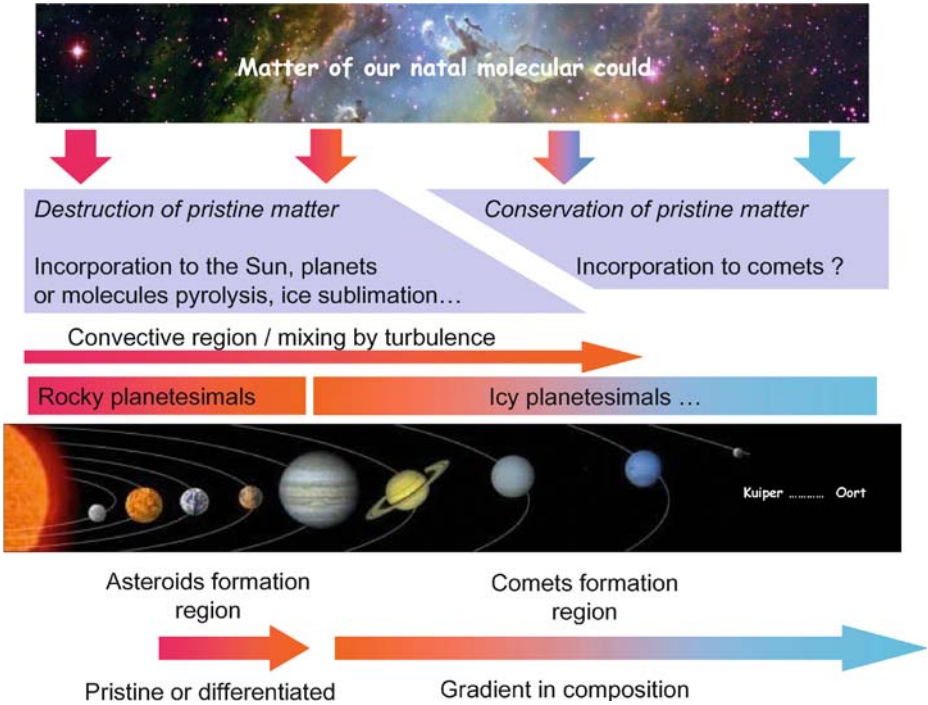


Fig. 13.3. Evolution of the matter between our natal molecular cloud, solar nebula and incorporation into the solar system bodies

grains close to the Sun, resulting in a loss of their initial composition. According to models, the turbulent region of the nebula can extend up to 30AU (Bockelée-Morvan et al. 2002; Hersant et al. 2001). Beyond, grains might have remained in a cold environment, and thus have kept an unaltered interstellar composition. Several models result from those considerations, in which the interstellar organic matter undergoes different levels of transformation before it is stored in comets. They are sometimes considered as simple unaltered interstellar grain aggregates (Greenberg 1982), or from a rather different point of view, made of matter completely re-processed in the solar nebula (Prinn and Fegley 1989). Other models consider an intermediate scenario (Lunine et al. 1991; Iro et al. 2003), which are probably more realistic (see Despois and Cottin 2005 for more details on this topic). However, putting aside the discussion about the origin of the cometary matter, observations show an undeniable abundance of a large variety of organic compounds. Molecules such as HCN, HCHO or HC₃N, which have been detected in several comets, can have an origin in the interstellar medium or inside the nebula but keep the same astrobiological interest. Moreover, in situ measurements in the close vicinity of comets 1P/Halley (in 1986) and 81P/Wild 2 (in 2004) bear witness to the existence of more complex structures. These later re-

main in the solid phase on dust grains when ejected from the nucleus, due to their high molecular mass (Huebner 1987; Kissel and Krueger 1987; Kissel et al. 2004; Mitchell et al. 1992). The lack of liquid water flowing in the nuclei over a long period of time does not allow to seriously consider the possibility of life on comets (even if short events of ice melting are possible for brief periods after the nucleus formation due to the radioactive decay of radioelements such as ^{26}Al (Podolak and Prrialnik 1997)). Still, all detected molecules, to which the complex molecules inferred from laboratory experimental simulations can be added, turn comets into objects with great astrobiological potential as endogenous sources for prebiotic molecules. Comets are the target of several past, current and future space missions, on which we will focus later in this chapter.

Nevertheless, molecules of astrobiological interest can be destroyed by the same processes that lead to their formation once synthesized in the interstellar medium or in the solar nebula (UV, charged particles or thermal cycles). Experiments conducted in space, onboard the MIR station, or space capsules FOTON, have shown that amino acids can survive exposition to solar UVs if they are embedded in a mineral matrix (on cometary grains for example, but also in meteorites) (Barbier et al. 1998; Barbier et al. 2002; Boillot et al. 2002). After surviving a long stay in space, those molecules have to endure the energy release when comets and meteorites impact the Earth. Theoretical (Chyba et al. 1990) and laboratory (Blank et al. 2001) works have shown that amino acids can survive such collisions.

Meteorites and more specifically those belonging to the carbonaceous chondrite family are another exogenous source for organic molecules. Unlike comets, for which no direct analysis of the nucleus composition has ever been made, a large number of meteorites have been studied with the most sensitive instruments in laboratories on Earth since by definition meteorites are bodies reaching Earth surface (the most famous among them being Murchison, Murray and Orgueil). The current flux of meteorites is estimated to be about 10 tons/year (Bland et al. 1996), and was probably higher on the primitive Earth. Great amounts of organic molecules have been detected in meteorites: hydrocarbons, alcohols, carboxylic acids, amines, amides, heterocycles including uracil, adenine and guanine (Stoks and Schwartz 1979; Stoks and Schwartz 1981), more than 70 amino acids for the Murchison meteorite, and recently diamino acids in the same object (Meierhenrich et al. 2004). Those detections show obviously that molecules once synthesized in space are able to survive impacts. Enantiomeric excess at the level of a few percents have been measured for some amino acids in the Murchison and Murray meteorites (Cronin and Pizzarello 1997; Pizzarello and Cronin 2000). This could give us a key to understanding the origin of homochirality in living organisms on Earth. Unlike comets, it is clear that parent bodies of meteorites have gone through events with liquid water, which might have lead to more advanced stages of chemical evolution. No space mission has been currently developed to analyze the composition of carbonaceous asteroids, which are probably the parent bodies of chondrites. Therefore, it is not in the

frame of this chapter to discuss further those objects. Our point is simply to underline the astrobiological importance of meteorites as an exogenous source of complex organic material. A very detailed review dealing with organic molecules in meteorites can be found in Botta and Bada (2002). However, a project to explore a carbonaceous asteroid could certainly be selected in the years to come, and this will give us important information about the evolution stage of those bodies and the origin of their organic component.

Micrometeorites are another vector for exogenous delivery of organic molecules. With an asteroidal or cometary origin, their current flux is estimated to be about 10 000 tons a year. They slowly sediment in the terrestrial atmosphere and thus undergo little warming that could destroy their organic content. Amino acids have been detected in micrometeorites collected in Antarctica (Maurette 1998). Therefore, they could also have played an important role in the origin of life process.

These three different kinds of exogenous delivery methods (comets, meteorites and micrometeorites) not only occurred on Earth, but also throughout the solar system (Fig. 13.1). However, those ingredients require liquid water to reach an increased level of complexity that could have lead to life.

13.1.1.2 The Endogenous Track

During the 1950s, for the first time, Stanley Miller implemented an experiment that consisted of simulating the chemistry coupling of a model of the primitive Earth atmosphere and the oceans. He submitted a gaseous mixture made of H_2 , CH_4 , NH_3 and H_2O to an electric discharge that simulated lightning storms, this mixture being connected with a warm liquid water bulb. This experiment resulted in the detection of a large amount of organic molecules, including several amino acids (Miller 1953). Those measurements are the basis of the chemical evolution theory, showing that the chemistry between simple and abundant molecules (in planetary atmosphere in the case of Miller's experiment) synthesizes key compounds leading to the formation of life, as we know it on Earth. The choice of a reduced atmosphere (C as CH_4 and N as NH_3) was motivated by the observations of the giant planets (detection of H_2 , CH_4 and NH_3), which are supposedly not to have evolved since their formation, and were therefore considered by Miller as good models for the primitive atmosphere of the telluric planets. Nevertheless, it is now considered that Earth's primitive atmosphere was dominated by CO_2 and N_2 , just like Venus and Mars. Such a composition does not allow syntheses of complex organic molecules, as shown in Table 13.1. However, if the Earth mantle was less oxidized 3.8 billions years ago than today, an important amount of methane could have been emitted through volcanism (Kasting 1993; Selsis and Parisot 2001). In this case, endogenous "Miller kind syntheses" were possible. However, to date, we do not have any indication of the amount of reduced gas in Earth's primitive atmosphere. Unlike exogenous deliveries that are still observed nowadays, there is no evidence that atmospheric endogenous syntheses actually occurred in the Earth's prebiotic environment.

Table 13.1. Organic molecules synthesized during “Miller-like” experiments as a function of the composition of the starting mixture (adapted from Raulin 2001). Compounds measured in the solid phase residues are usually detected after an acid hydrolysis

Gaseous mixture	Related planetary atmosphere	Organic products	
		Electric discharge	Photolysis
CH ₄ + NH ₃ + H ₂ O (+ H ₂)	Giant planets	RH (saturated & unsaturated) HCN & other nitriles (saturated), RCO ₂ H, H ₂ CO other aldehydes Ketones & alcohols <i>Solid:</i> Amino acids & nitrogenated heterocycles after hydrolysis	RH (mostly sat) HCN RCN (saturated) if N/C < 1 RNH ₂ if N/C > 1 H ₂ CO, other aldehydes Ketones & alcohols <i>Solid:</i> Amino acids after hydrolysis
CH ₄ + N ₂ (+ H ₂ O)	Titan, Triton	RH (saturated & unsaturated) HCN & other nitriles (saturated & unsaturated) including HC ₃ N, other aldehydes and C ₂ N ₂ , H ₂ CO Ketones & alcohols <i>Solid:</i> Amino acids & nitrogenated heterocycles after hydrolysis	RH (saturated & unsaturated) H ₂ CO & other aldehydes with low yields <i>Solid:</i> Carboxylic acids
CO + NH ₃ + H ₂ O		HCN, oxygenated organic compounds <i>Solid:</i> Amino acids after hydrolysis	
CO ₂ + N ₂ + H ₂ O + CO/H ₂	Primitive Earth?	RH (mostly saturated) HCN, other nitriles (saturated) H ₂ CO, other aldehydes, ketones <i>Solid:</i> Amino acids after hydrolysis	
CO ₂ + N ₂ + H ₂ O	Primitive Earth? Venus, Mars	No synthesis	

However, unlike the Earth, Titan, the largest moon of Saturn, has a noticeable fraction of methane in addition to an atmosphere mainly composed of nitrogen. As shown in Table 13.1, organic syntheses are very efficient in such an environment, which is confirmed by observations since we can observe a thick layer of organic aerosols resulting from N_2/CH_4 complex chemistry. Yet, the temperature of the satellite does not allow liquid water on the surface. Therefore, Titan can be considered as a laboratory at the planetary scale in which we can study the level of complexity that chemical evolution can reach without any water in the liquid phase. Clearly, there is no serious reason to consider that life might have appeared on Titan. Nevertheless, the presence of oceans is possible under a deep ice layer (Fortes 2000), but transport between surface organics and underground liquid water is still an open issue. Titan is the destination of the European probe Huygens, which will be later developed in this chapter.

Finally, another kind of endogenous source of organic compounds has to be considered: a synthesis at the bottom of the oceans, in hydrothermal vents (also known as black smokers) (Corliss et al. 1981). When oceanic plates are drifting apart, water infiltrates the crust and springs from black smokers at high temperatures, enriched in gas (H_2 , N_2 , CH_4 , H_2S , etc.) and minerals. It has been shown experimentally that amino acids can be synthesized in such conditions (high temperature and pressure, reduced environment, and minerals which can act as a catalyst) (Hennet et al. 1992; Yanagawa and Kobayashi 1992). However, such molecules are also very efficiently destroyed due to the very high temperatures in those environments. More experimental data and field measurements are required to assess the feasibility of this mechanism.

13.1.2 Follow the Water

Water in the liquid state seems to be the most favorable solvent to allow molecules with an exogenous or endogenous origin to reach more evolved structures, and thus to make possible the appearance of life. If the H_2O molecule is ubiquitous in the solar system and the galaxy, it is most of the time in the gaseous phase (Venus, Mars, giant planets, interstellar medium) or the solid phase (giant planets icy moons, comets, interstellar ices). To date, the presence of water in the liquid state has only been established on the Earth. Yet, two solar system objects are of prime interest: Mars and Europa.

13.1.2.1 Mars

NASA's current strategy for Martian exploration is summarized in the expression: "follow the water". Even if nowadays there is no evidence of liquid water on the surface of the red planet, the record of past aqueous flows can be read in the landscapes morphology, and in the composition of some minerals or rocks, which can only be formed in the presence of liquid water. A large number of space missions to Mars slowly lift the veil on its past. As we will see later in

this chapter (see also Chap. 4), it is now quite well-established that an abundant amount of water once flowed on Mars (rivers, or even an ocean covering almost the whole northern hemisphere), and this, at a time period during which life had possibly already appeared on Earth. Thus, a source of organic molecules (the Martian primitive atmosphere being probably similar to the Earth's, the exogenous source hypothesis is the most likely) combined with liquid water, could have induced the appearance of life just like on Earth. However, attempts to detect organic molecules on Mars have not yet been successful. Today, those molecules and possible liquid water pockets, if any, are probably buried underground, and future space missions will have to dig the oxidized soil of the planet to add a new dimension to our knowledge of Mars. If life developed on Mars, independently from the Earth², this could mean that the jump from chemistry to biology is written in the laws of the *natural* evolution of organic matter each time the requirements "organic matter plus liquid water" are fulfilled.

13.1.2.2 Europa and the Icy Moons of the Giant Planets

Observations of the Jovian satellite Europa from the space probe Galileo lead to an accumulation of data that converge to the hypothesis that an ocean is buried under its icy surface (morphological clues (Carr et al. 1998), detection of a magnetic field (Kivelson et al. 2000), and hydrated salts at the surface (McCord et al. 1999)). The ocean could have been created by the heating of the planet due to Jupiter's gravitational field, which puts Europa out of shape and induces important stresses on the ice. It could be hundreds of kilometers deep and would be covered with an icy layer of thickness 10 to 50km (Fig. 13.4) (Sotin et al. 2002), which casts the feasibility of a submarine exploration mission into the rather distant future. Nevertheless, external exploration missions are currently considered to confirm the presence of an ocean under the ices of Europa: JIMO (Jupiter Icy Moons Orbiter) was an example of project for such a mission.³

However, the association between liquid water and a source of organic molecules is still problematic on Europa. According to the thickness of the ice layer, exchanges between exogenous deliveries at the surface and the ocean are possible but could be very limited (Pierazzo and Chyba 2002; Schenk 2002). Concerning endogenous sources, in the absence of an atmosphere, organic syntheses in the vicinity of black smokers could be the only way to synthesize prebiotic compounds. Indeed, theoretical models have shown that the putative ocean of Europa could be in contact with the silicate mantle of the planet. If an inner

² It is also possible that life appeared on only one of the two planets, and then was exported to the other: Martian meteorites have indeed been collected on Earth showing that such a travel is feasible, one of these meteorites contains very controversial hints of past microbial activity: ALH 84001. Experiments implemented in space have shown that some microorganisms can survive such an interplanetary journey.

³ For more information about JIMO:

http://en.wikipedia.org/wiki/Jupiter_Icy_Moons_Orbiter

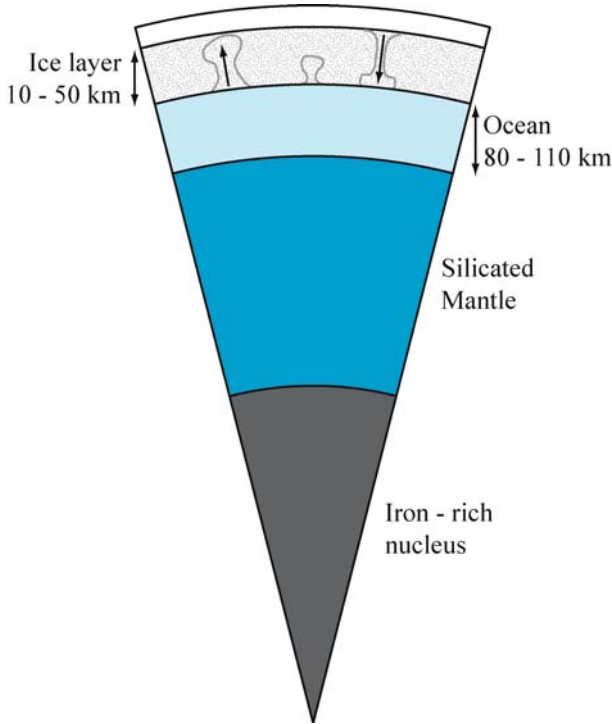


Fig. 13.4. A representation of the internal structure of Europa (Sotin et al. 2002)

source of heat from inside the satellite does exist, hydrothermal vents could be present at the bottom of the ocean, even if these considerations are still very speculative. Thus, a conjunction of those favorable conditions could make Europa another object of the solar system currently harboring life.

The presence of oceans is also considered inside Ganymede (McCord et al. 2001) and Callisto (Zimmer et al. 2000). However, in both cases, they would be embedded at larger depths than on Europa, and trapped between two ice layers, where the presence of hydrothermal vents is not possible. Ganymede and Callisto are also targets of the JIMO mission.

13.2 The Space Exploration of Comets

13.2.1 General Considerations

As already mentioned in this chapter, comets, as are planets, are made of matter from a molecular cloud, which collapsed to give birth to our solar system. Those icy planetesimals have not been through differentiation processes thanks to their small size; they have been stored in the outermost, hence coldest, regions of the

solar system (Kuiper Belt and Oort Cloud). Therefore their initial composition should remain unchanged and thus bear testimony of the physical and chemical environment prevailing in the solar nebula region in which they have been formed. Some might even kept the same composition as our natal molecular cloud (Fig. 13.3). Therefore, comets are considered as the oldest archives of our planetary system, as if samples of pristine matter were stored in a freezer to be preserved. A gravitational perturbation or a collision can eject a cometary nucleus out of its reservoir and place it on an elliptical orbit that will lead it close to the Sun. This approach triggers the sublimation of the ices (first CO, then H₂O and other volatile compounds which are frozen far from the Sun) and leads to the formation of the cometary atmosphere (the coma) and of the tails that can sometimes be seen with naked eyes from the Earth (Fig. 13.5) (much more about comets can be found in Despois and Cottin 2005).

Most of the currently available data about the composition of comets are derived from observations by teledetection methods from Earth. To date, about twenty molecules have been detected in the gaseous phase (mainly H₂O, but also CO, CO₂, CH₃OH, NH₃, HCN, CH₄, etc.), but there is no direct information about the molecular composition of cometary nuclei (Bockelée-Morvan et al. in press; Despois and Cottin 2005). Those observations are the starting point for a large number of laboratory experiments, which consist of simulating the behavior of ices made of molecules detected in the coma. Those ices are

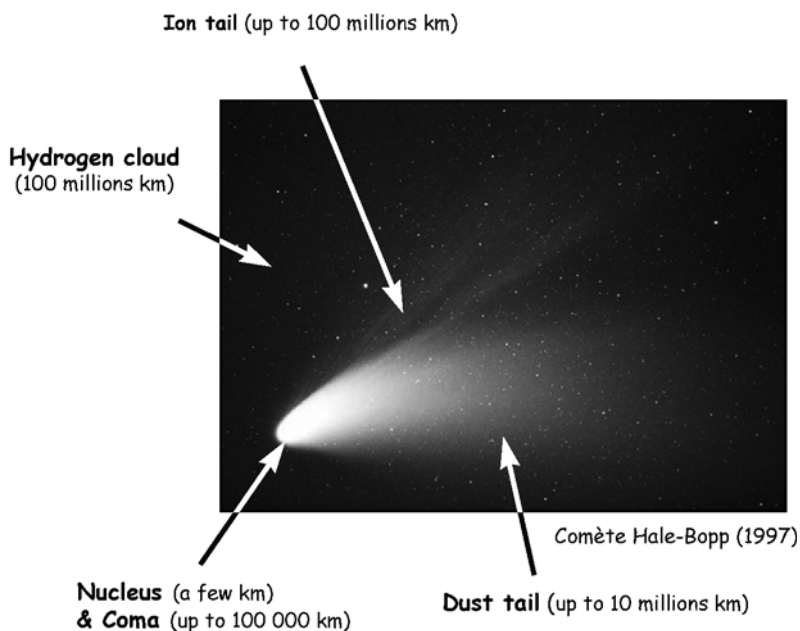


Fig. 13.5. Structure of a comet at about 1 AU from the Sun

submitted to different kinds of energy sources (UV photons, charged particles, thermal cycles) simulating the conditions they encountered during their history (where history can go back to long before being incorporated into comets, maybe even into the interstellar medium). Those experiments yield the production of molecules much more complex than the starting material, and are a sign for great complexity of the nucleus in terms of organic composition (Table 13.2).

Table 13.2. Molecules detected during experimental simulations of cometary and interstellar ice analogues. *Italic letters* refer to molecules actually detected in comets, (*t*) means tentative detection only in the analogues. Amino acids (alanine, AIB, and so on, except glycine) were detected after acid hydrolysis of the room temperature residue (from Despois and Cottin 2005)

Hydrocarbons:

CH₄
C₂H₂, C₂H₄, C₂H₆
C₃H₈, C₄H₁₀
C₅H₁₀, C₅H₁₂
C₆H₁₂, C₆H₁₄
C₇H₁₆

Amides:

NH₂CHO
CH₃CONH₂
HOCH₂CONH₂
NH₂(CO)₂NH₂
HOCH₂CH(OH)CONH₂

Amines:

HOCH₂CH₂NH₂
HCNH(NH₂)
 Diaminopyrrole
 Diaminofurane
 Triaminopropane
(CH₂)₆N₄ (HMT)

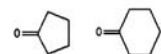
Aldehydes:

H₂CO
CH₃OCH₂CHO (t)

Others: *CO, CO₂, C₃O₂, H₂O₂, H₂CO₃, N₂H₄, HNCO, NH₂CONH₂, NH₂CONHCONH₂*

Ketones:

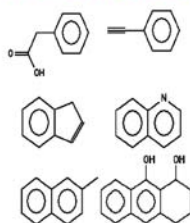
CH₃COCH₃
HOCH₂COCH₃
HOCH₂CH₂COCH₃



Carboxylic acids:

HCOOH
CH₃COOH (t)
HOCH₂COOH
HOCH₂CH(OH)COOH
HOCH₂CH₂COOH
NH₂COCO₂H

Aromatic Compounds:



Ethers:

CH₃OCH₂OCH₃ (t)
C₃H₆O₃ (Trioxane) (t)
(-CH₂-O-)_n (POM)

Alcohols:

CH₃OH
CH₃CH₂OH
HOCH₂CH₂OH
HOCH₂CH(OH)CH₂OH
C₄H₈(OH)₂
C₅H₉OH (t)
C₅H₁₁OH

Amino Acids:

NH₂CH₂COOH (Glycine)
NH₂CH(CH₃)COOH (Alanine)
CH₃CH₂CH(NH₂)COOH (α ABA)
CH₃CH(NH₂)CH₂COOH (β ABA)
(CH₂NH₂)(CH₃)CHCOOH (AIBA)
 Sarcosine
 Ethylglycine
 Valine, Proline, Serine
 Aspartic acid
 Diaminopropanoic acid
 Diaminobutyric acid
 Diaminopentanoic acid
 Diaminohexanoic acid

Esters:

HCOOCH₃
CH₃COOCH₃
CH₃CH₂COOCH₃

Space missions are a necessary complement to observation from the Earth and laboratory experimental simulations.

Finally, since the nucleus is hidden behind the coma, its observation from the Earth is impossible. If the comet is not active, the nucleus is much too far and too dark to be observed. Thus, we have the pictures of only four cometary nuclei, which have been taken from space probes (Fig. 13.6).

13.2.2 Past Missions

Past cometary space missions are presented in Table 13.3 and Fig. 13.7. It is not possible to summarize all the results obtained from those missions in the frame of this chapter; therefore, we have chosen to focus more specifically on astrobiology related data.

Organic compounds with molecular weights higher than those observed by teledetection methods (from the Earth or from space probe instruments) have

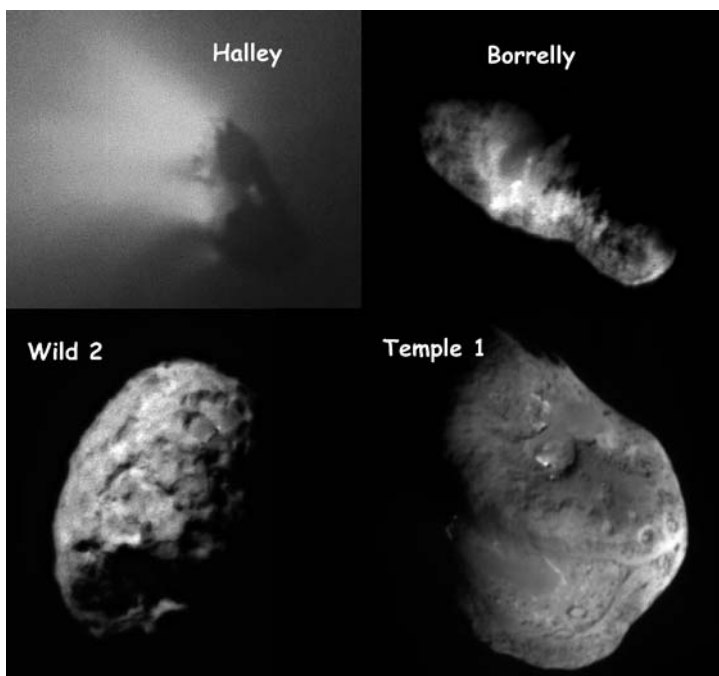


Fig. 13.6. The four cometary nuclei observed to date: *Halley* from the Giotto spacecraft in 1986 (distance at picture time: ~ 700 km, nucleus: $4.1 \times 4.2 \times 8$ km) (courtesy of ESA/MPIfA); *Borrelly* from the Deep Space 1 spacecraft in 2001 (distance: 3417 km, nucleus: 3.2×8 km) (courtesy of NASA/JPL); *Wild 2* from Stardust in 2004 (distance: 500 km, nucleus: ~ 5 km) (courtesy of NASA/JPL); *Temple 1* from Deep Impact in 2005 (distance: 3000 km, nucleus: $\sim 5 \times 7$ km) (courtesy of NASA/JPL-Caltech/UMD)

Table 13.3. Specifications and main results for past cometary space missions

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
ISEE-3/ICE Mass: 390 kg Nominal power: 173 W	NASA	08/12/1978: Launch 1981: End of nominal mission	The International Sun-Earth Explorer 3 was part of a program of three probes conceived to study interactions between the Sun and Earth magnetosphere. The probe was then renamed International Cometary Explorer and headed toward comet Giacobini-Zinner. The probe will perhaps be captured in August 2014 and then exhibited in a space museum	Instruments for measurements of magnetic fields, plasma, ions, X and γ -rays	First confirmation that comets are made of ice and dust as predicted by Whipple (1950) Closest Giacobini-Zinner approach: 26 650 km
SAKIGAKE Mass: 138.1 kg Nominal power: 100 W	ISAS (Japan)	01/07/1985: Launch 03/11/1986: Comet Halley flyby	This mission is the first space probe launched independently from the USA and USSR It was mainly consisting in a feasibility study for Suisei probe, both missions being devoted to Comet Halley	Mainly magnetic fields and solar wind measurement instruments	Was used as a distant reference point for the interpretation of the results of other probes that got closer to the nucleus Closest Halley approach: 7×10^6 km
Means “Pioneer” in Japanese					

Table 13.3. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
SUISEI Mass: 139.5 kg Nominal power: 100W Means “comet” in Japanese	ISAS	08/18/1985: Launch 03/08/1986: Comet Halley flyby	Twin probe of SAKIKAGE, except for the instrument After Halley flyby, attempts to redirect the spacecraft towards comets Giacobini-Zinner and Temple-Tuttle failed because of lack of fuel	Mainly a UV imagery camera and a solar wind measurement instrument	Study of the hydrogen cloud around the nucleus of Comet Halley Closest Halley approach: 152 400 km Was hit by at least two cometary grains (about 1mm in diameter) at this distance
VEGA 1 Mass: 4920 kg (including 1500 kg for the Venusian module) Vega is a contraction between Venera (name of the soviet Venusian missions) and	USSR	12/15/1984: Launch 06/11/1985: Arrival and release of the Venusian probe	Twin probe of Vega 2, it had two goals: to release an entry probe in the atmosphere of Venus (atmospheric balloon), and the flyby of Comet Halley. Concerning comets, it was conceived to measure the physical parameters of the nucleus (size, shape, temperature and surface properties), the structure	1 – TVS: Television system, 2 – TKS: 3 channels spectrometer, 3 – IKS: Infra-red spectrometer, 4 – PUMA: Mass spectrometer for dust grains, 5 – SP-1 & 2: Dust particles counter, 6 – ING: Mass spectrometer for neutral gas, 7 – PLASMAG:	Vega 1 & 2 probes were the first to accomplish Comet Halley flyby. Pictures taken at that time were used to refine the final approach of Giotto spacecraft. On Vega 1 pictures two bright regions could be distin-

Table 13.3. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
Gallei (Halley in Russian)			and dynamic of the coma, the composition of gas and grains, and solar wind interactions	Plasma energy analysis, 8 – TUNDE-M: Energetic particles analysis, 9 – MISCHA: Magnetometer, 10 – APV-N & V: Waves and plasma analysis, 11 – DUCMA: Dust particles detector, 12 – MSU-TASPD: Energetic particles analysis	guished, and were first interpreted as two distinct nuclei (whereas it was latter established that it was indeed more active regions on the nucleus). First detection of organic molecules on grains. Closest Halley approach: 8890 km
VEGA 2	USSR	12/21/1984: Launch 15/06/1985: Arrival and release of the Venusian probe	See VEGA 1	See VEGA 1	Closest Halley approach: 8030 km
Mass: 4920 kg (including 1500 kg for the Venusian module)		03/09/1986: Comet Halley flyby			Speed relative to nucleus: 78 km/s

Table 13.3. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
GIOTTO Mass: 582.7 kg	ESA	07/02/1985: Launch	First probe of the European Space Agency. Its main goals were: (1) to collect color pictures of the nucleus, (2) to determine the molecular and isotopic abundances in the coma, (3) to characterize the physical and chemical processes in the coma and in the ionosphere, (4) to determine the molecular and isotopic composition of cometary grains, (5) to measure the gas and dust production, and (7) to study comets/solar wind interactions. After Halley flyby, even if some instruments were damaged (including the camera), the probe was redirected towards comet Grigg-Skjellerup	1 – HMC: Halley Multicolor Camera (nucleus pictures); 2 – MAG, Magnetometer, 3 – DID: dust impact detection system (dust flux and mass measurements), 4 – RPA: Retarding Plasma Analyser which consists in ESSA: electrostatic analysis + PICCA: Positive Ion Cluster Composition Analyser (mass spectrometer), 5 – JPA: Johnstone Plasma Analyser (two instruments devoted to the study of the solar wind and charged particles), 6 – OPE: Optical Probe,	First picture of a cometary nucleus. With a very low albedo, Comet Halley nucleus is one of the darkest objects of the solar system. The probe detected complex organic matter on grains. Halley closest approach: 596 km (14 seconds before closest approach, a large impact interrupted transmissions during 32 minutes) Speed relative to nucleus: 68 km/s. After Halley flyby, the spacecraft was
Nominal power: 196 W		03/14/1986: Comet Halley flyby			
The name					
Giotto was chosen because the artist painted a comet on one of his paintings		07/10/1992: comet Grigg-Skjellerup flyby			

Table 13.3. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
				Experiment (brightness of the coma) 7 – EPA: Energetic Particles (electrons, protons and α -particles measurement), 8 – GRE: Giotto Radio Experiment (electronic environment), 9 – PIA: Particulate Impact Analyser, 10 – NMS: Neutral Mass Spectrometer, 11 – IMS: Ion Mass Spectrometer (9, 10, 11: three mass spectrometers for the analysis of gas and grains compositions)	covered in at least 26 kg of dust. During the encounter with comet Grigg-Skjellerup (at less than 200 km from the nucleus), Giotto was able to measure surprising variations in the magnetic field, and gathered evidences that a second nucleus might be present in the coma
DEEP SPACE 1	NASA	10/24/1998: Launch	Technological mission mainly devoted to testing ionic propulsion.	1 – MICAS: Miniature Integrated Camera Spectrometer (UV, Visible, IR imager),	Second cometary nucleus to be photographed. No water ice detected
Mass: 374 kg					
Nominal power: 2500 W		07/29/1999: Asteroid 9969	First objective was the		

Table 13.3. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
		Braille flyby 09/22/2001: comet Borrelly flyby	flyby of an asteroid Then, the mission was extended to the flyby of two comets: Wilson-Harrington in January 2001 and Borrelly in September 2001. Due to a failure of the navigation system, only comet Borelly en- counter was feasible	2 – PEPE: Plasma Experiment for Planetary Exploration, 3 – FGM: Fluxgate magnetometer	at the surface Asteroid Braille closest approach: 26km, relative speed: 15.5km/s comet Borrelly closest approach: 2171 km, relative speed: 16.6km/s
CONTOUR Mass: 328 kg	NASA	07/03/2002: Launch 08/15/2002: Contact lost	The goal of the mission was a detailed study of three comets: Encke, Schwassman- Wachmann 3, and comet d'Arrest or another one according to new detections during the mission. The contact with the probe was lost	1 – CRISP: CONTOUR Remote Imager/Spectro- graph, (Visible and infrared imager), 2 – CFI: CONTOUR Forward Imager (Imager and navigation), 3 – NGIMS: Neutral Gas Ion Mass Spectro- meter,	Mission lost

Table 13.3. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
			during the ignition of one of the engines. It is thought that the temperature increase caused the spacecraft destruction	4 – CIDA: CONTOUR Dust Analyzer (Mass spectrometer for dust analysis)	
			Missions not dedicated to comets, but with related opportunity results		
GALILEO	NASA	10/18/1990: Launch	Direct observation of the collision between comet Shoemaker-Levy 9 and Jupiter (July 16 to 22, 1994), which took place on the other side of Jupiter when seen from the Earth		
Exploration of Jupiter and its moons		09/21/2003: Impact with Jupiter			
ULYSSE	ESA	06/10/1990: Launch	Crossed by chance the tails of comet Hyakutake, and showed that cometary tails were much longer than expected		
Sun exploration and more specifically its poles (out of the ecliptic plane)		05/1996: Encounter with comet Hyakutake tails			



Fig. 13.7. Past cometary missions (courtesy of NASA)

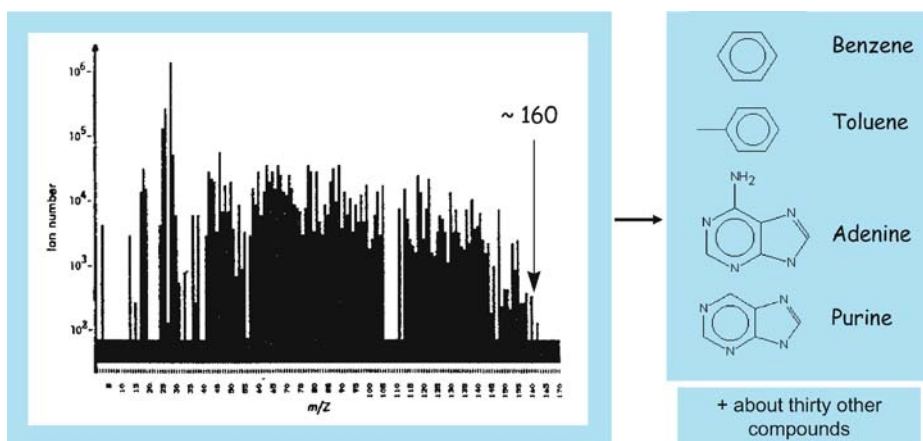


Fig. 13.8. Accumulated mass spectra collected in Comet Halley by the instrument PUMA on board Vega 1 spacecraft and tentative interpretation (adapted from Kissel and Krueger 1987)

been detected by mass spectrometry, mainly with the PICCA instrument (part of RPA) onboard the Giotto spacecraft, and the PUMA instrument, onboard Vega 1. The first was dedicated to the analysis of the gaseous phase, whereas the latest was conceived to analyze the composition of cometary dust grains. The most fruitful related studies have been reported by Kissel and Krueger (1987). To interpret the PUMA mass spectra, they made hypotheses about the sudden energy dissipation processes near solid surfaces after an impact to infer the rules of molecular ion formation. Such impacts occur when a dust particle hits the mass spectrometer's target. This hypothesis yields a good agreement be-

tween their predictions and other observations for small molecules (e.g., HCN, CH₃CN). For larger molecules, the problem is more complex, as they cannot survive the impact process without fragmentation. In addition to the difficulty of assembling correctly the fragments to reconstitute the original molecule, the nature of the fragments themselves is uncertain, several fragments having the same mass (at the spectrometer's resolution). Thus the identification by the authors of purines and adenine, for example, should be considered as very tentative. However, what is clearly shown by the PUMA mass spectra is that compounds with complex structure and large molecular masses (at least with m/z up to $160\text{g} \times \text{mol}^{-1}$) are present in grains, which agrees with results from laboratory simulations.

On the other hand, first interpretations of the heavy-ion mass spectra from the PICCA instrument led Huebner (1987) to infer the presence of polyoxymethylene (POM, the formaldehyde polymer) in the Comet Halley. The alternation of patterns with a ratio $m/z = 14$ and 16 is in good agreement with the succession of $-\text{CH}_2-$ ($m = 14$) and $-\text{O}-$ ($m = 16$) that form the polymer (Fig. 13.9). Huebner's hypothesis seemed to be confirmed by laboratory studies that measure POM mass spectra, which are in agreement with observations (Möller and Jackson 1990; Moore and Tanabe 1990). However, Mitchell et al. (1992) showed that the mass spectrum pattern which led Huebner to announce the detection of polyoxymethylene is only characteristic of a mixture of molecules composed of C, H, O and N atoms (CHON molecules). Therefore, PICCA mass spectra do not imply the presence of POM, but of a large diversity of CHON molecules, among which POM could be present. Moreover, measurements with the NMS (Neutral Mass Spectrometer) instrument, onboard Giotto, showed that

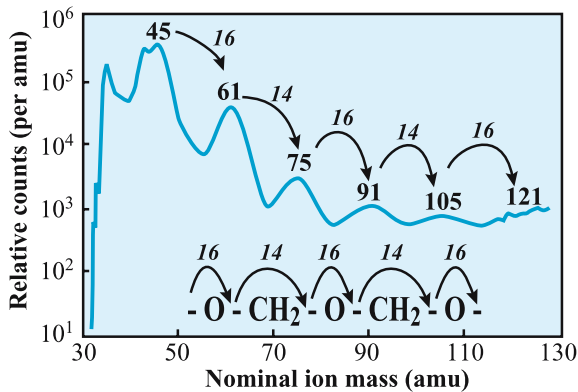


Fig. 13.9. Mass spectrum obtained with the PICCA instrument on board Giotto, March 14, 1986. This is a mean of spectra acquired between 8200 and 12 600 km from the nucleus. (Huebner 1987)). Peaks near $m/z = 45$ are saturated, and masses beyond 120 are dominated by the apparatus noise. Comparison of the results with POM structure

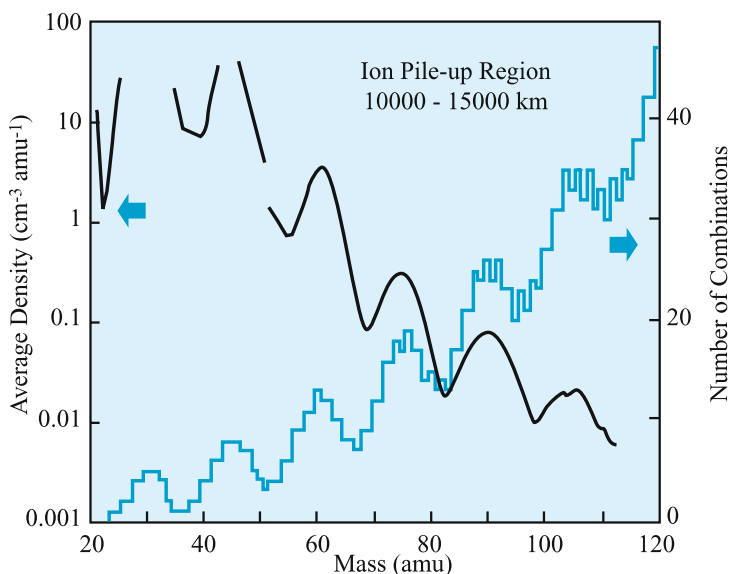


Fig. 13.10. PICCA mass spectrum compared to the number of combinations between C, H, O, N atoms leading to a molecular ion that can be observed (Mitchell et al. 1992)

the formaldehyde in Comet Halley could be produced by an extended source⁴ throughout the coma of Comet Halley (Meier et al. 1993). It has recently been shown that such observations can be explained by the presence of POM on grains (a few percents in mass). The polymer would be slowly decomposed into H_2CO as it is heated and photolyzed (Cottin et al. 2004). Without being final evidence for the presence of POM in comets, this polymer is so far the best explanation to this extended source. Once it is imported to Earth's primitive oceans, it could be an interesting source of concentrated formaldehyde, and play a key role in the synthesis of prebiotic sugars. Table 13.4 shows all the molecules detected with mass spectrometry in Halley's vicinity. Those results, even if they are not always very specific, show that comets are extremely rich in terms of organic composition, and thus are of prime interest for astrobiology studies. Current and future missions have to take up the challenge of formal identification of complex molecules in comets.

13.2.3 Current Missions

In the infrared and radio wavelength domains, simple organic molecules have very specific signatures, and thus, they can be detected from the Earth by remote

⁴ The H_2CO spatial distribution in the coma is not compatible with an emission from the nucleus only. This supports the hypothesis that there is an (at least one) additional source which produces formaldehyde as it spreads outwards from the nucleus. This is called an extended source.

Table 13.4. Organic molecules inferred from mass spectra of gas and dust particles in Comet Halley. The confidence levels of detections by MS are established as follows: *Confirmed*: molecule also detected by remote observations; *High*: molecule not detected by remote observations but present after laboratory irradiations of cometary ice analogues; *Medium*: molecule detected only by mass spectroscopy with a good confidence level according to the authors; *Low*: molecule only inferred by mass spectroscopy with a low confidence level according to the authors

Molecule	Family	Mass spectrometer	Confidence level	Reference
Hydrocyanic acid	C-N-H	PUMA, PICCA	Confirmed	1, 2, 3
Methyl cyanide	C-N-H	IMS	Confirmed	7
Acetonitrile	C-N-H	PUMA, PICCA	Confirmed	1, 2, 3
Methanol	C-O-H	NMS	Confirmed	6
Formaldehyde	C-O-H	PUMA, PICCA	Confirmed	3
Formic acid	C-O-H	PUMA	Confirmed	1, 2
Acetaldehyde	C-O-H	PUMA, PICCA	Confirmed	1, 3
Ammonia	N-H	PUMA	Confirmed	2
Isocyanic acid	C-N-O-H	PUMA	Confirmed	1, 2
Ethane	C-H	NMS	Confirmed	8
Acetylene	C-H	NMS	Confirmed	8
Acetic acid	C-O-H	PUMA	High	1
Polyoxymethylene	C-O-H	PICCA	High (itself or derivatives)	4, 5
Ethene	C-H	NMS	High	8
Iminoethane	C-N-H	PUMA	Medium	1, 2
Aminoethene	C-N-H	PUMA	Medium	1, 2
Pyrroline	C-N-H	PUMA	Medium	1
Pyrrole	C-N-H	PUMA	Medium	1, 2
Imidazole	C-N-H	PUMA	Medium	1
Pyridine	C-N-H	PUMA	Medium	1, 2
Pyrimidine	C-N-H	PUMA	Medium	1, 2
Ethyl cyanide	C-N-H	IMS	Medium	7
Pentyne	C-H	PUMA	Low	1
Hexyne	C-H	PUMA	Low	1
Butadiene	C-H	PUMA	Low	1
Pentadiene	C-H	PUMA	Low	1
Cyclopentene	C-H	PUMA	Low	1
Cyclopentadiene	C-H	PUMA	Low	1
Cyclohexene	C-H	PUMA	Low	1
Cyclohexadiene	C-H	PUMA	Low	1

Table 13.4. (continued)

Molecule	Family	Mass Spectrometer	Confidence Level	Reference
Benzene	C-H	PUMA	Low	1
Toluene	C-H	PUMA	Low	1
Propanenitrile	C-N-H	PUMA	Low	1
Iminomethane	C-N-H	PUMA	Low	1
Iminopropene	C-N-H	PUMA	Low	1
Purine	C-N-H	PUMA	Low	1, 2
Adenine	C-N-H	PUMA	Low	1, 2
Polyaminocyanomethylene	C-N-H	PICCA	Low	5
Methanolitrile	C-N-O-H	PUMA	Low	1
Methanalimine	C-N-O-H	PUMA	Low	1
Aminomethanol	C-N-O-H	PUMA	Low	2
Aminomethanal	C-N-O-H	PUMA	Low	2
Oxyimidazole	C-N-O-H	PUMA	Low	1
Oxypyrimidine	C-N-O-H	PUMA	Low	1
Xanthine	C-N-O-H	PUMA	Low	1

1 (Kissel and Krueger 1987); 2 (Krueger and Kissel 1987); 3 (Krueger et al. 1991); 4 (Huebner 1987); 5 (Huebner et al. 1989); 6 (Eberhardt and Krankowsky 1995); 7 (Geiss et al. 1999); 8 (Altwegg et al. 1999)

sensing. This is not the case for complex compounds, for which low abundances and/or less specific spectroscopic signatures requires other investigation means. This can be illustrated by the three current space missions dealing with comets: Stardust (sample return from comet Wild 2), Deep Impact (impact with comet Temple 1 nucleus) and Rosetta (aiming to intercept comet 67P/Churyumov-Gerasimenko, and for which one of the instruments is a gas chromatograph).

Information about those missions is presented in Table 13.5 and Fig. 13.11. Stardust mission consists in two phases: collecting cometary dust grains with an aerogel, which is now completed, then the analysis of the grains in Earth laboratories after recovering the samples in January 2006. As far as the analysis of the molecular composition of grains and their astrobiological relevance are concerned, most of the science part will be done after the sample returns. Yet, it is interesting to note that evidences of the presence of complex organic material on grains have already been obtained thanks to the mass spectrometer onboard Stardust (CIDA instrument) (Kissel et al. 2004). However, those analyses have been performed on grains that may have evolved since their release from the

Table 13.5. Specifications and main results of current space missions

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
STARDUST Mass: 300 kg	NASA	02/07/1999: Launch	First mission with comet sample return: dust grains trapped in an aerogel during a comet flyby.	1 – Imagery camera, 2 – Aerogel for grains collection, 3 – CIDA: Cometary and Interstellar Dust Analyzer (mass spectrometer), 4 – DFMI: Dust Flux Monitor Instrument	Third cometary nucleus to be photographed, after Halley and Borrelly. Pictures have shown a surprising topography: craters (probably due to the outgassing of the nucleus, which cause the surface to collapse), cliffs and pits.
Nominal Power: 330 W		02–05/2000: First interstellar grains collection	Once brought back to Earth, their physical properties and chemical composition can be analyzed with the most recent and sensitive instruments. The very low density of the aerogel should allow a progressive deceleration of dust grains, in order to minimize any alteration by heating, and pyrolysis of organic molecules. Comet Wild 2		In situ analysis of dust grains by mass spectrometry corroborate previous measurements on Comet Halley grains, and confirm
		07–12/2002: Second interstellar grains collection			
		01/02/2004: Comet Wild 2 flyby and cometary grains collection			
		01/15/2006: Earth sample return	is a relatively new comet in the inner solar system, which passed only a few times in Sun's vicinity		

Table 13.5. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
DEEP IMPACT Mass: 650 kg + 370 kg (impactor)	NASA	01/12/2005: Launch 07/04/2005: Impact with comet Temple 1	The purpose of the mission is to know and characterize the inside of a comet. Its main feature is an impactor that was launched toward comet Temple 1 to excavate a crater on its surface. During the impact the science objectives are: (1) observe how the crater forms,	On spacecraft: 1 – HRI: High Resolution Instrument (visible and infrared) 2 – MRI: Medium Resolution Instrument (visible, mainly for navigation during the last 10 days of the approach) 3 – Impactor (370 kg, impact planned at about 10 km/s)	the presence of complex organics. Several thousand grains have been collected and will be brought back to Earth early 2006. Wild 2 closest approach: 236 km Successful impact. Fourth cometary nucleus to be photographed. Closest picture: 30 km from nucleus, with 4 m resolution. Presence of impact craters (therefore not similar to previous observed nucleus). Diversity in appearance:
Nominal power: 92 W					

Table 13.5. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
			(2) measure its depth and diameter, (3) measure the composition of the interior of the crater and the ejectas, and (4) determine the changes in outgassing after the impact. On impact, the crater produced is expected to range in size from that of a house to that of a football stadium (~ 115 m), and two to fourteen stories deep (more than 30 m)	On impactor: 4 – ITS: Impactor Targeting Sensor (Similar to MRI, visible imagery to direct the impactor and produce nucleus pictures just before the impact)	smooth and cratered regions
ROSETTA	ESA	03/02/2004: Launch	First mission whose goal is to <i>land</i> on a cometary nucleus. It was first conceived to explore comet Wirtanen, but because of the failures of Ariane 5 launcher	ROSETTA: Observations: OSIRIS: Optical, Spectroscopic, and Infrared Remote Imaging System,	Instruments
Mass: 1200 kg					
Nominal power: 850 W		03/2005: First Earth gravity assist			
The name of the mission refers to the					

Table 13.5. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments
Rosetta stone that gave Champollion a crucial clue to understand Egyptian hieroglyphs. The lander was called Philae following the name of the island on which an obelisk was found. Its association with the Rosetta stone helped Champollion in his discovery. Comets are considered as the Rosetta		02/2007: Mars gravity assist 11/2007: Second Earth gravity assist	the mission was postponed and a new target chosen: 67P/Churyumov-Gerasimenko. The science objectives of the mission are: (1) a global characterization of the nucleus (dynamic properties, surface morphology), and composition (2) determination of the chemical, mineralogical and isotopic compositions of volatiles and refractories in a nucleus, as well the links between the gaseous and the solid phase, (3) study of the development of cometary activity and the processes in the surface layer of the	ALICE: UV spectrometer, VIRTIS: Visible and Infrared Thermal Imaging Spectrometer, MIRO: Microwave Instrument for the Rosetta Orbiter Composition analysis: ROSINA: Rosetta Orbiter Spectrometer for Ion and Neutral Analysis, COSIMA: Cometary Secondary Ion Mass spectrometer, MIDAS: Micro-Imaging Dust Analysis System Physical properties of the nucleus and the coma: CONSERT: Comet Nucleus Sounding ExperimentT, GIADA: Grain Impact Analyzer and Dust Accumulator, RPC: Rosetta orbiter Plasma Consortium, RSI: Radio Science

Table 13.5. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments
stones that will help us in understanding the origin of the solar system		08/2014: orbital insertion with comet Churyumov-Gerasimenko 11/2014: Philae delivery	nucleus and the inner coma (dust/gas interaction), (4) a global characterization of asteroids (dynamic properties, surface morphology and composition)	PHILAE: APX: Alpha-p-X rays spectrometer, COSAC: Cometary Sampling and Composition experiment, MODULUS: Method Of Determining and Understanding Light elements from Unequivocal Stable isotope compositions, CIVA/ROLIS: Rosetta Lander Imaging System, SESAME: Surface Electrical, Seismic and Acoustic Monitoring Experiments, MUPUS: Multi-Purpose Sensor for surface and subsurface science, ROMAP: Rosetta Magnetometer and Plasma monitor, CONSERT: Comet Nucleus Sounding

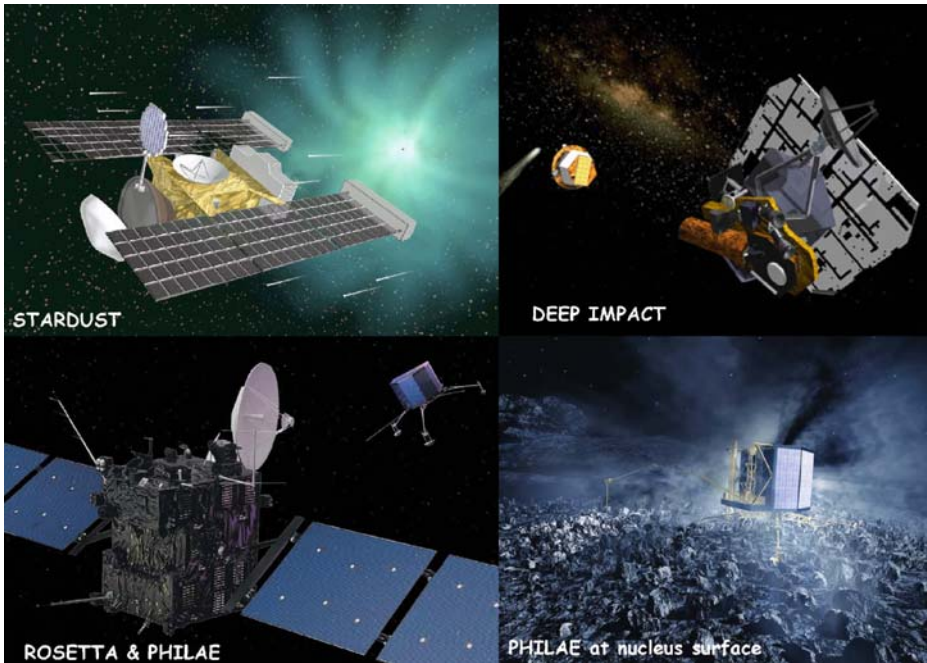


Fig. 13.11. Current space missions (courtesy of NASA & ESA)

nucleus, and for the grains that will be analyzed on Earth, one cannot rule out some alteration when impacting the aerogel. Therefore, measurements directly on the nucleus are necessary.

Deep Impact mission aims to gain access to the interior of a comet, and then observe what could be the most pristine matter of the solar system, since cometary inside layers have been protected from irradiation and heating due to previous Sun approaches. A collision between an impactor and comet Temple 1 was planned on July 4, 2005, in order to excavate a large crater at the surface. The impact was a real success as can be seen in Fig. 13.12. At the time this chapter was written, only imagery data has been released, therefore we cannot discuss further the results in terms of nucleus properties and chemical composition. It has only been shown that there will be almost no change in the cometary trajectory because of the Deep Impact collision. New chemical species may be detected during the increase of outgassing after the impact, either with the on-board infrared instrument, or with other remote sensing instruments on Earth. Many new data concerning comets are expected when the results of this mission become published.

As for the Rosetta mission, it consists of two parts: one orbiter revolving around the nucleus, and a lander called Philae. The instruments onboard the orbiter will enable an unprecedented analysis of the composition of volatile and



Fig. 13.12. Cometary nucleus 90 seconds before the impact seen from the impactor at 900 km (*on the left*) and impact event seen from the spacecraft (*on the right*). (courtesy of NASA/JPL-Caltech/UMD)

refractory compounds released from the nucleus. Three instruments (ROSINA, COSIMA and MIDAS) will collect and analyze gas and dust as close as 1 km from the surface of the nucleus, it is expected that they will be close enough to study almost unaltered matter compared to its release from the nucleus. Nevertheless, probably the most fruitful information from an astrobiological point of view will come from Philae. In this chapter, we have decided to emphasize the COSAC instrument since most of the new organic molecules should be detected thanks to this experiment.

The COSAC instrument is a gas chromatograph (GC) coupled with a mass spectrometer (MS, a linear time of flight spectrometer in this case). It consists of 8 chromatographic columns, each of them being connected to its own detector (TCD), but it is also possible to connect them to a mass spectrometer (Rosenbauer et al. 1999) (Fig. 13.13). Previous results obtained thanks to direct mass spectrometry measurements with Puma, Giotto and Stardust spacecrafts, gave *only* the mass spectrum of the mixture of all the molecules at the same time. COSAC will carry out a preliminary separation by chromatography, which will achieve a quasidefinite identification of the compounds since they will be identified both from their retention time⁵, and from their individual mass spectra. Samples will be collected after drilling the surface and heated at various temperatures before being injected into the analysis system. Pyrolysis (up to 600°C) is possible in order to degrade the most refractory component, and then enable gas phase analysis of the fragments. Out of the eight chromatographic

⁵ The time for a compound to be eluted from the chromatographic column, this time depends on the properties of both the molecules and the stationary phase of the column.

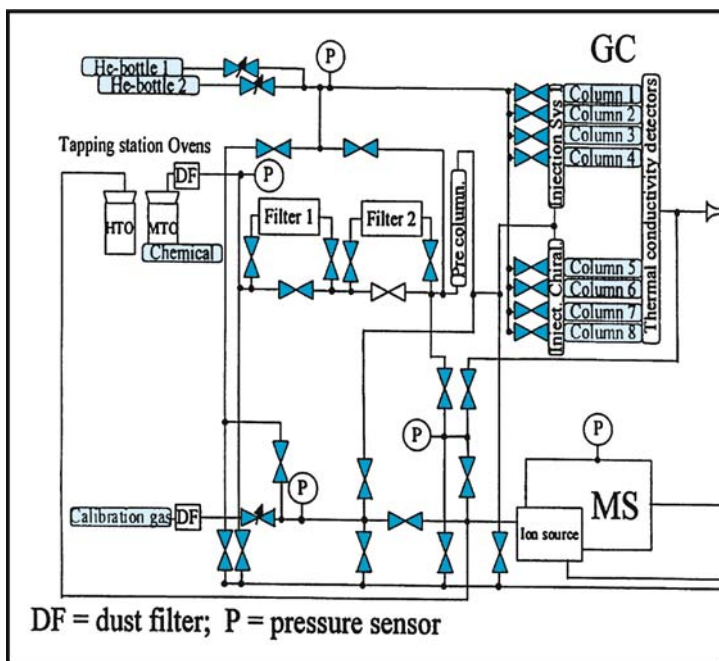


Fig. 13.13. The COSAC experiment (Rosenbauer et al. 1999)

columns, three are specifically devoted to the analysis of chiral molecules in order to distinguish enantiomers. The other five columns have been selected so that a maximum of molecules can be detected. Moreover, the simultaneous analysis of a single sample with several columns (up to four columns at the same time) will facilitate the data analysis by comparison, and thus increase the interpretation's reliability. Table 13.6 shows all the molecules that were considered for the selection of the chromatographic columns, and the ones that will be actually possible to detect. One can note that amino acids and other heavy compounds such as oxalic acid, urea, etc., are not detectable, since they are not volatile enough to be analyzed in the gaseous phase. GC analysis of such compounds requires a preliminary stage called derivatization (chemical reaction that makes the targeted compound more volatile). This procedure is not feasible with the COSAC instrument, but work is in progress to include derivatization in future Martian exploration experiments (Rodier et al. 2001). Nucleus analyses by the COSAC instrument will be completed by CIVA (infrared analysis) and MODULUS (for isotopic measurements).

It is interesting to note that HMT, a molecule that is very often produced during laboratory simulations of cometary ices (Bernstein et al. 1995; Cottin et al. 2001; Muñoz Caro and Schutte 2003), will be detected by the COSAC instrument if it is present on the nucleus. Indeed, even if this compound is

Table 13.6. Organic molecules that might be detected in comets thanks to the COSAC instrument (Szopa et al. 2003)

List of the compounds considered for the study (Abbreviations: d: detected in comets; s: generated during experimental simulations; a: analysable by the COSAC GC subsystem within the space conditions; p: potentially; italic: compounds not vaporisable without decomposition)



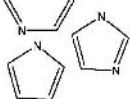
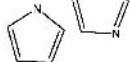











Compounds	Formula	d	s	a	Compounds	Formula	d	s	a
Hydrocarbons									
Methane	CH ₄	x	x	x	Alcohols				
Ethane	C ₂ H ₆	x	x	x	Methanol	CH ₃ OH	x	x	x
Ethene	C ₂ H ₄	x	x	x	Ethanol	C ₂ H ₅ OH	x	x	
Acetylene	C ₂ H ₂	x	x	x	Propan-2-ol	C ₃ H ₇ OH			x
Allene	H ₂ C=C=CH ₂	x	x	x	2-Me-Propan-1-ol	(CH ₃) ₂ C ₂ H ₅ OH		x	x
Propane	C ₃ H ₈	x	x	x	Butan-1-ol	C ₄ H ₉ OH			x
Butane	C ₄ H ₁₀	x	x	x	Pentan-1-ol	C ₅ H ₁₁ OH		x	x
Pentane	C ₅ H ₁₂	x	x	x	1,4-Butanediol	HOC ₄ H ₈ OH		x	x
Cyclopentane	Cyclo-C ₅ H ₁₀	x	x	x	Ethylene glycol	HOC ₂ H ₄ OH		x	x
3-Me-Pentane	(C ₂ H ₅) ₂ CHCH ₃	x	x		Glycerol	(CH ₂ OH) ₂ CHOH		x	x
Hexane	C ₆ H ₁₄	x	x		Ketones				
Heptane	C ₇ H ₁₆	x	x		Acetone	CH ₃ COCH ₃		x	x
Octane	C ₈ H ₁₈	x	x		Hydroxypropanone	HOCH ₂ COCH ₃		x	x
Norane	C ₉ H ₂₀	x	x		3-Hydroxybutan-2-one	HOC ₂ H ₄ COCH ₃		x	
Decane	C ₁₀ H ₂₂	x	x		Cyclopentanone	Cyclo-C ₅ H ₈ CO		x	x
					Cyclohexanone	Cyclo-C ₆ H ₁₀ CO		x	x
Aldehydes									
Formaldehyde	H ₂ CO	x	x	x	Esters				
Acetaldehyde	CH ₃ CHO	x			Methyl formate	HCOCH ₃	x	x	x
					Methyl acetate	CH ₃ COCH ₃		x	x
Ethers/oxides									
Dimethylether	CH ₃ OCH ₃		x	x	Methyl propionate	C ₂ H ₅ COCH ₃		x	x
Diethylether	C ₂ H ₅ OC ₂ H ₅		x	x	Ethyl acetate	CH ₃ COC ₂ H ₅			x
Dimethoxymethane	CH ₃ OCH ₂ OCH ₃		x	x	Rare gases and inorganics				
Trioxane			p	x	Helium	He			
					Neon	Ne			x
Amides									
Formamide	HCONH ₂	x	x	x	Argon	Ar		p	x
Acetamide	CH ₃ CONH ₂		x	x	Krypton	Kr			x
Urea	(NH ₂) ₂ CO		x		Xenon	Xe			x
Biurea	H ₂ NCONNCONH ₂		x		Carbon monoxide	CO		x	x
Hydroxyacetamide	HOCH ₂ CONH ₂		x		Carbon dioxide	CO ₂		x	x
Oxamide	NH ₂ COCONH ₂		x		Ammonia	NH ₃		x	x
Glyceramide	HOCH ₂ CH(OH)CONH ₂		x		Nitrogen	N ₂		x	x
Carboxylic acids									
Formic acid	HCOOH	x	x	x	Amines				
Acetic acid	CH ₃ COOH		x	x	Methanamine	CH ₃ NH ₂			
Oxamic acid	H ₂ NCOOOH		x		Ethaneamine	C ₂ H ₅ NH ₂			
3-Hydroxypropionic acid	HOC ₂ H ₄ COOH		x		2-Ethanolamine	NH ₂ C ₂ H ₄ OH		x	x
					Formamide	HCNH(NH ₂)		x	
					Pyridine			p	x
					Imidazole			p	x
					Pyrrrole			p	x
Glycolic acid									
Glycolic acid	HOCH ₂ COOH		x		Sulfur species				
Glyceric acid									
Glyceric acid	HOCH ₂ CH(OH)COOH		x		Disulfur	S ₂		x	
					Sulfur dioxide	SO ₂		x	x
					Hydrogensulfide	H ₂ S		x	x
					Thioformaldehyde	HC(S)H		x	
					Carbonylsulfide	OCS		x	x
					Carbon disulfide	CS ₂		x	x
					Sulfur monoxide	SO		x	
Nitriles									
Cyanic acid	HCN	x		x	Sulfur species				
Isoyanic acid	HNC		x		Disulfur	S ₂		x	
Acetonitrile	CH ₃ CN		x	x	Sulfur dioxide	SO ₂		x	x
Acrylonitrile	CH ₂ =CHCN		x	x	Hydrogensulfide	H ₂ S		x	x
Methacrylonitrile	CH ₂ C(CH ₃)CN			x	Thioformaldehyde	HC(S)H		x	
Propanenitrile	C ₂ H ₅ CN		p		Carbonylsulfide	OCS		x	x
Butanenitrile	C ₃ H ₇ CN			x	Carbon disulfide	CS ₂		x	x
Isobutanenitrile	(CH ₃) ₂ CHCN			x	Sulfur monoxide	SO		x	

Table 13.6. (continued)

Compounds	Formula	d	s	a
Cis-crotonitrile	cis-CH ₃ C ₂ H ₂ CN			x
Trans-crotonitrile	trans-CH ₃ C ₂ H ₂ CN			x
Amino acids				
<i>Glycine</i>	NH ₂ CH ₂ COOH		x	
<i>Alanine</i>	CH ₃ CH(NH ₂)COOH		x	
<i>α-ABA</i>	C ₂ H ₅ CH(NH ₂)COOH		x	
<i>β-ABA</i>	CH ₃ CH(NH ₂)CH ₂ COOH		x	
<i>AIBA</i>	(CH ₂ NH ₂)(CH ₃)CHCOOH		x	

Aromatics

Benzene		p		x
Toluene		p		x
Phenylacetylene			x	x
Benzeneacetic acid			x	x
Indene			x	x
Quinoline			x	x
2-Me-Naphthalene			x	x
Naphthalene				x
Anthracene				
Phenanthrene				
<i>Diol(tetrahydro)-anthracene</i>			x	

Compounds	Formula	d	s	a
Others				
Polyoxymethylene	(-CH ₂ -O-)n	p	x	
Hydrazine	N ₂ H ₄		x	
Carbonic acid	H ₂ CO ₃		x	
Hydrogen Peroxide	H ₂ O ₂		x	
Carbon suboxide	OC ₃ O		x	
Isocyanic acid	HNCO	x	x	
Water	H ₂ O	x	x	x
Hexamethylenetetramine			x	x



easily synthesized in the conditions simulated in the laboratory, it remains undetected in comets or interstellar medium. HMT could be a source of amino acids once imported to Earth, since its hydrolysis yields the formation of amino acids (Wolman et al. 1971). Its actual detection in comets would validate the contribution of laboratory experimental simulation to understanding the chemistry of cometary ices. On the other hand, COSAC will not be able to detect POM (see Sect. 13.2.2) because it is readily degraded into formaldehyde when heated. Infrared measurements might be sufficient to detect it, but the spectroscopic signatures of this compound could be overlapped by other high molecular weight organic molecules.

Thus, our knowledge in terms of molecular composition of comets will do a great leap forward at the time of Rosetta's arrival near comet 67P/Churyumov-

Table 13.7. Some data about comet 67P/Churyumov–Gerasimenko

Discovery: 1969 by Klim Churyumov and Svetlana Gerasimenko
Orbital period: 6.57 yrs
Aphelion: 858 million km, i.e., 5.73 AU
Perihelion: 194 million km, i.e., 1.3 AU
Mean nucleus diameter: 1980 m

Gerasimenko. With the new detected molecules, we will be in a better position to evaluate the parameters necessary to their production, and then trace back the physicochemical condition prevailing in the solar system when the comets were accreted, or even better, to the composition of our natal molecular cloud. Therefore, the astrobiology relevance of the comet will be disproved or strengthened.

13.2.4 Future Space Missions

The next step of comets exploration will probably be a return of samples taken at the surface of a nucleus (and underneath). There is not yet such a mission under development, but proposals have already been submitted to NASA and nucleus sample return missions are to be expected to be accepted in the years to come.

As already mentioned earlier in this chapter, knowing the composition of one comet will not allow us to extrapolate the results to the whole family of comets, as their composition should depend on the place they were formed (see Fig. 13.3). Then, one shall vary our targets selections (from the Kuiper Belt, the Oort Cloud, few or many perihelion passages), before pretending to know these mysterious objects.

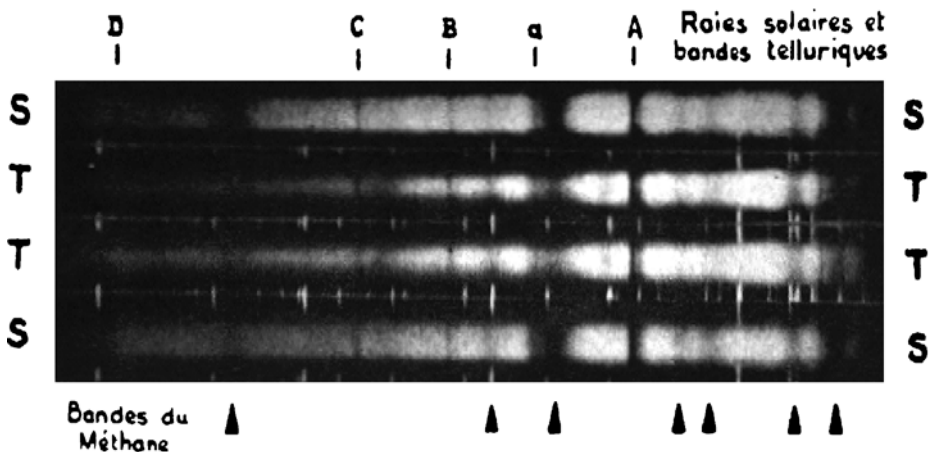
13.3 The Space Exploration of Titan

13.3.1 Observations and Models of Titan Before Space Missions

On the night of March 25, 1655, using a refracting telescope with a focal length of almost four meters and a magnifying power of 50, Christiaan Huygens observed for the first time a small star aligned with the rings of Saturn at a distance of three arc seconds.⁶ During the following days, he made new observations and noted the displacement of this object as compared to fixed stars and Saturn. This star followed Saturn's course in the sky, and thus Huygens deduced that it must be a satellite. He had just discovered the moon Titan.

In 1944, making spectroscopic measurements of Titan in the atmospheric window between 550 and 850 nm, Kuiper observed for the first time the methane

⁶ Angular distance projected on the sky.



Band	Wavelength (Å)	Origin
A	7594	O ₂
a	7187	H ₂ O
B	6867	O ₂
C	6563	H α
D	5890,5896	Na

Fig. 13.14. Saturn (*S*) and Titan (*T*) spectra obtained by Kuiper, at the MacDonalld observatory (1944) showing characteristic methane absorption bands between 5500 and 9000 Å. The wavelength calibration is given by the Fraunhofer⁷ and telluric absorption bands

absorption bands (Fig. 13.14). In 1965, Low (1965) made the first measurements of the flux emitted from Titan in the 10µm infrared atmospheric window. He determined a brightness temperature⁸ of ~160K. However, it is only in the beginning of the 1970s that the development of space missions and infrared detectors allowed the discovery of the extraordinary rich organic chemistry occurring in the atmosphere of this satellite.

From infrared observations obtained in the 8–14µm window, Gillett et al. (1973) attributed the emissions bands to the ν_4 methane band and the ν_7 ethane band. At the same time, Trafton (1972a,b) detected traces of molecular hydrogen. On their side, Veverka (1973) and Zellner (1973) showed, analyzing polarimetric observational data, the presence of a dense and highly diffusive layer in the atmosphere. Finally, Caldwell (1975) using the OAO-2 (Orbital Astronomical Observatory) satellite, obtained photometric measurements in the ultraviolet wavelength range, around 260nm and inferred a very low albedo, on the order of 0.05.

⁷ Absorption lines visible in the solar spectra.

⁸ Temperature of an object considered as a black body.

The infrared band of methane and ethane are observed in emission, which can be explained by a temperature inversion in the atmosphere. Titan's atmospheric models have this inversion together with the high brightness temperature, the low ultraviolet albedo and the polarimetric measurements. A first set of models, developed initially by Pollack (1973), included a dense cloudy highly diffusive layer that could explain the visible spectrum and polarimetric data. A greenhouse effect that could be produced by induced absorption of a 0.5 bar equimolar H_2 and CH_4 mixture, would explain the 160 K temperature. Nevertheless, this model does not account for the low ultraviolet albedo and the temperature inversion. Consequently, a second type of model had to be constructed. Since the Rayleigh diffusion (diffusion by the gas) cannot explain the low albedo, Danielson et al. (1973) proposed that small particles ("dust"), efficiently absorbing ultraviolet photons, could be present in the high atmosphere. This hypothesis is supported by the fact that at the same period Khare and Sagan (1973) produced in their simulation experiment (photon irradiation of a $\text{NH}_3 - \text{CH}_4 - \text{H}_2$ mixture) a brown polymer that had similar properties compared to Danielson dust. Thus, if such material would be produced in the high atmosphere of Titan, these aerosols would have a very small size and could not emit efficiently the absorbed photons. Consequently, their temperature would increase until their emission equilibrates the absorbed energy. The temperature of the gas would then increase through collision with those aerosols up to the observed 160 K. In this model, the surface would have a black body temperature of 80 K.

The two models are very different but a compromise seems possible. Hunten (1978) proposed a model in which the low atmosphere is described by the first model and the high atmosphere by the second model. Taking into account the low gravity on Titan⁹, the hydrogen amount proposed by Pollack (1973) is not realistic; it is therefore replaced by molecular nitrogen. The choice of N_2 comes from the studies about the initial composition of Titan. Nitrogen could have been trapped in the ices in the form of ammonia that could then produce molecular nitrogen photolysis. Atreya et al. (1978) calculated that, if Titan had an important greenhouse effect at the beginning of its history ($T > 150\text{K}$), NH_3 could have led to the formation of 20 bars of N_2 . Hunten's model includes N_2 as the main constituent (with 0.25% CH_4 and 0.5% H_2), a cloud with a highly diffusing upper part, and a high atmosphere containing dust. In this model, the tropopause is determined by the altitude where the optical depth (N_2 pressure induces absorption opacity plus cloud opacity) equals one. This region of the atmosphere is supposed to have an effective temperature of 77 K and a pressure of ~ 600 mbar. The surface pressure is then deduced from the adiabatic approximation from the tropopause down to the ground. This leads to a surface pressure of ~ 20 bars for a surface temperature of 200 K compatible with the millimeter observations (Conklin et al. 1977). Nevertheless, radio observations in the centimeter domain where the atmosphere is supposed to be transparent, lead to brightness temperature of 100 K. The main uncertainty on this measurement

⁹ Low gravity favors light element such as hydrogen to escape from the atmosphere.

comes from the poor angular resolution, which could lead to confusion between Titan and Saturn. Still, as pointed out by Hunten, his model could also lead to a lower temperature: for example, a 100K surface temperature would lead to a 2-bar surface pressure.

Finally, one can note that the photochemical model developed by Strobel (1974) with an atmosphere mainly composed of CH_4 and H_2 in same quantities (this model was initially developed for Jupiter with $\text{CH}_4/\text{H}_2 \ll 1$), predicts the production of C_2H_6 , C_2H_2 , and in smaller amounts C_2H_4 , and C_3H_4 as trace constituents. The presence of C_2H_2 and C_2H_4 has been later confirmed by observational data obtained by Gillett (1975) in the 8 to 14 μm wavelength range. CH_3D was also detected at that time.

13.3.2 Voyager Missions at Titan

Launched in 1977, the Voyager 1 and 2 spacecrafts arrived in the Saturn system three years later after a flyby of Jupiter in 1979, where they obtained a large amount of data. The Saturn encounter was planned so that the maximum information could be gathered about the rings, the satellites and the magnetosphere, and to allow the study of Saturn's and Titan's atmospheres. Titan studies focused on several major goals: determine the diameter of the solid surface, the temperature and pressure profiles, the atmospheric composition and search for a magnetic field (Stone and Miner 1981).

Voyager 1 spacecraft obtained the largest amount of data on Titan. In fact, its trajectory was planned to approach the satellite as much as it could and its closest approach was at less than 7000km from the center of Titan (Ness et al. 1981). A first disappointment came from the pictures obtained (Smith et al. 1981). In fact, even if the presence of aerosol was expected, some scientists were hoping that some hole in the haze would allow for the observation of the surface. It was not the case. A uniform orange haze was entirely hiding the surface. Nevertheless, the images showed three distinct aerosol layers (Fig. 13.15). The first one was observed around a 200km altitude. It was dense, optically thick in the visible and seemed to be made of particles having a $\sim 0.3\mu\text{m}$ radius at the top. A second layer detached from the first one (even if a small amount of aerosols was observed in between) was seen 100km above and was about 50km thick. The last layer was observed at a mean altitude of 400km and was also observed by the ultraviolet occultation experiment. Finally, the observations showed a north-south asymmetry, the north being much darker than the southern hemisphere.

The greatest progress in our knowledge of this satellite of Saturn came from radio and ultraviolet occultation measurements and from infrared and ultraviolet experiments. The radio occultation experiment (~ 3.6 and 13cm) sounded the atmosphere (Tyler et al. 1981) down to the surface allowing the determination of the radius of $2575.0 \pm 0.5\text{km}$ (Lindal et al. 1983). This type of sounding measures the variation of the refractive index of the atmosphere as a function of altitude. Since the refractivity is proportional to the density, this allows to determine

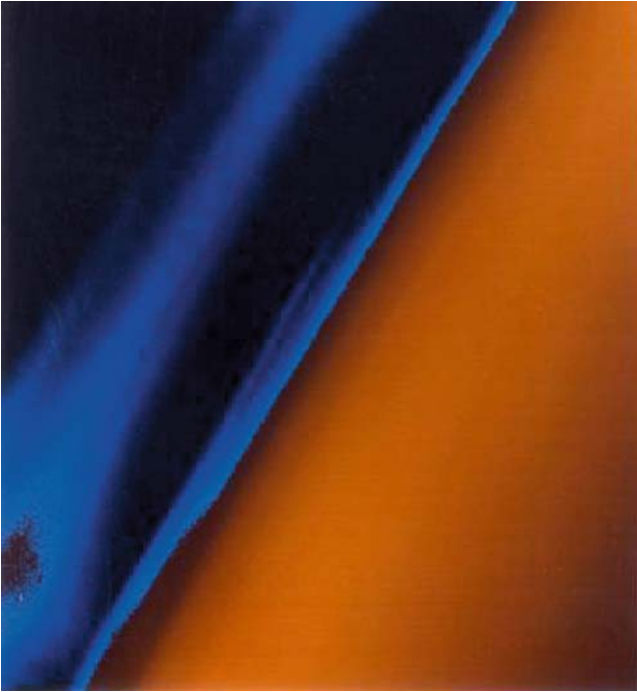


Fig. 13.15. Aerosol layers observed by Voyager 1 (courtesy of NASA)

the density profile of the atmosphere. Assuming hydrostatic equilibrium, the pressure profile can be deduced. Finally, the choice of a state equation allows to retrieve the ratio $T/\langle m \rangle$ as a function of altitude (T is the temperature and $\langle m \rangle$ the mean molecular mass).

Assuming that the atmosphere is entirely composed of molecular nitrogen, Lindal et al. (1983) determined the density, pressure and temperature profiles of Titan atmosphere. The temperature at the ground level is 94K and the pressure 1495mbar. The temperature decreases in the troposphere and the profile shows the presence of a tropopause with a temperature of 71.2K at 40km altitude ($p \sim 120$ mbar). The temperature then increases in the stratosphere to reach a temperature of ~ 170 K at 200km ($p \sim 0.75$ mbar) (see Fig. 13.18). The choice of molecular nitrogen as the major constituent seemed to be arbitrary since methane was the only abundant detected component. However, it has been a posteriori justified by the results obtained with the infrared and ultraviolet instruments.

The ultraviolet spectroscopic experiment (UVS for UltraViolet Spectrometer), which did measurements in the 60 to 160nm wavelength range, observed in addition to Lyman α emission due to atomic hydrogen, an intense emission around 100nm (Broadfoot et al. 1981). This emission corresponds to the one

obtained in laboratory experiments when N_2 is excited by electronic impacts. The relative intensity of the bands allows to determine a column density of $\sim 10^{15} \text{ cm}^2$. Besides, the ultraviolet occultation experiment which measured the attenuation of the solar flux through the atmosphere allowed to measure the atmospheric scale length¹⁰ at an altitude of 1265km (altitude where the optical depth¹¹ $\tau = 1$ for wavelengths lower than 80nm), characteristic of nitrogen absorption: $H = 85 \pm 10 \text{ km}$ (Smith et al. 1982). Assuming that the atmosphere is entirely composed of nitrogen, this lead to a temperature $T = 176 \pm 20 \text{ K}$. Furthermore, knowing the absorption coefficient of N_2 around 70nm, one can determine the local density: $[N_2] = 2.7 \pm 0.2 \times 10^8 \text{ cm}^{-3}$. A change in the scale length starts at an altitude of 1125km and is attributed to methane. Smith et al. (1982) deduced from those measurements a mixing ratio $[CH_4]/[N_2] = 0.08 \pm 0.03$ (making the assumption of a constant temperature). They also determined a mixing ratio for acetylene of 1 to 2% above 825km and less than 0.3% below 675km. This would imply a high sink for this compound in this region of the atmosphere. The data also showed highly absorbing layers around 390km and 760km, with a $\sim 70 \text{ km}$ thickness. The first layer seemed to be well correlated to the high altitude layer observed in the visible while the second one is not. Since in this region, C_2H_2 abundance decreases rapidly one can think that it could be efficiently photolysed in this region of the atmosphere, leading to the formation of higher weight compounds. This emission spectrum has lead to the determination of upper limits for the mixing ratio of H_2 , Ar and Ne in the high atmosphere: $[H_2]/[N_2] \leq 0.06$, $[Ar]/[N_2] \leq 0.06$, $[Ne]/[N_2] \leq 0.01$ and $[CO]/[N_2] \leq 0.05$ (Strobel and Shemansky 1982). As a conclusion, the UV data showed that nitrogen is the major compound in Titan's atmosphere, so that it should be the place for an active photochemistry involving nitrogen and methane. The confirmation came from the analysis of the infrared spectroscopic measurements.

The infrared spectrometer (IRIS for Infrared Radiometer and Interferometer Spectrometer) onboard the Voyager spacecraft was a Fourier transform spectrometer (resolution 4.3 cm^{-1}) covering the domain 180 to 2400 cm^{-1} (55 to $4.2 \mu\text{m}$). The spectrometer of the Voyager 1 spacecraft took almost 3000 spectra during the Titan flyby and a wide latitudinal-longitudinal geographical region was explored. The first analysis of the spectra obtained at the equator confirmed the presence of methane (CH_4), ethane (C_2H_6), acetylene (C_2H_2) but also the presence of ethylene (C_2H_4) and hydrogen cyanide (HCN) (Hanel et al. 1981). All these compounds are seen in emission showing that they are present in the relatively warm stratosphere of Titan. The study of Titan's north pole completes this list adding propane (C_3H_8), methyl acetylene (C_3H_4) (Maguire et al. 1981), diacetylene (C_4H_2), cyanoacetylene (HC_3N), and cyanogen (C_2N_2)

¹⁰ The atmospheric scale length is the distance over which the pressure is divided by a factor e (2.718). It is proportional to the temperature, and inversely proportional to the mean molecular mass and gravity.

¹¹ The optical depth is the quantity, which represents the opacity of a material. When it equals one, the light intensity is attenuated by a factor e .

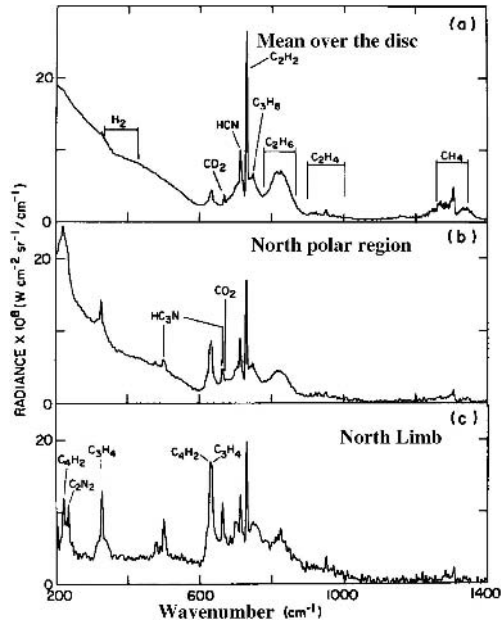


Fig. 13.16. Infrared spectra obtained at different latitudes (a and b) and at the limb (c) by the IRIS experiment of the Voyager 1 mission. The main detected compounds are quoted (from Samuelson et al. 1983)

(Kunde et al. 1981). The presence of all those compounds in the stratosphere confirm the complex and active chemistry that is occurring in the N_2/CH_4 mixture which composes Titan's atmosphere.

A more careful study of the spectra obtained at different latitudes allowed Samuelson et al. (1983) to discover the presence of an oxygenated compound: CO_2 (Fig. 13.16). In fact, even if its mole fraction is very low $\sim 1.4 \times 10^{-8}$, its distribution is uniform over Titan's disc, which is not the case for the other minor compounds. This first detection of oxygenated compound in Titan's atmosphere was later confirmed by the detection of CO in the near-infrared (Lutz et al. 1983) and millimeter (Marten et al. 1988) observations.

13.3.3 Similarities and Differences Between Titan and the Earth

The results obtained from the data accumulated by the Voyager missions have lead to a real breakthrough in our knowledge of Titan. One of the major results was the discovery of a dense atmosphere mainly composed of nitrogen in which the very active coupled chemistry of nitrogen and methane leads to the formation of numerous organic compounds. This makes Titan a unique object in the solar system since it the only one, outside the Earth, to have a dense atmosphere mainly made of nitrogen. Its slightly reductive atmosphere, in between the highly

Table 13.8. Past missions (Fig. 13.17)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
PIONEER 11 Mass: 259kg	NASA	04/05/1973: Launch	Pioneer 11 spacecraft was the second spatial vehicle to visit the external solar system but the first to explore Saturn	Instruments to measure magnetic field, UV photometer, IR radiometer and imaging photopolarimeter	Confirmed the presence of submicron aerosols having an increasing size with decreasing altitudes. Observed an hydrogen torus around Saturn
Nominal power: 165 W		09/01/1979: Saturn Flyby 11/1995: end of the mission			Observed an hydrogen torus around Saturn
VOYAGER 1 and 2 Mass: 722kg	NASA	V1 05/09/1977, V2 20/08/1977: Launch	These two missions succeeded to study in a small time interval all the giant planets and their satellites.	Imaging instruments, IR and UV spectrometry, magnetic and charge particles measurements	Established that N ₂ is the major gas. Measured the temperature and pressure profiles. Detections of numerous organics mainly hydrocarbons and nitriles. Closest approach to Titan:
Nominal power: 420 W		V1 11/1980, V2 08/1981: Saturn Flyby	The trajectory of Voyager 1 was optimized to do a close flyby of Titan		V1: 6490 km V2: 665 960 km



Fig. 13.17. Pioneer 11 and Voyager spacecrafts (courtesy of NASA)

reductive atmospheres of the giant planets and the highly oxidant atmospheres of the telluric planets, makes it a photochemical reactor which leads to the production of complex organic compounds. With this unique characteristic, Titan can be considered as a planetary scale laboratory for the study of the evolution of organic matter in abiotic conditions.

Titan's environment can help us to understand the primitive Earth and the chemical evolution that led to the emergence of Life. The analogy between both objects is underlined by several common aspects. Both have a dense atmosphere mainly composed of nitrogen and in which the temperature profiles present similar structures. Hence, if CO_2 is the major greenhouse gas in the terrestrial troposphere, in Titan, nitrogen, methane and hydrogen play this role through their pressure-induced absorptions. Warming in Titan's stratosphere is explained by ultraviolet absorption by small particles, which cannot emit efficiently in the infrared, where on Earth, ozone plays this role. The temperature in the terrestrial mesosphere is controlled by CO_2 infrared re-emission where on Titan it is mainly HCN, which plays this cooling role. Finally, in the very high atmosphere of both objects the gas is heated by ionization. Therefore, the structures of both atmospheres are similar because they are controlled by similar physical processes. Only the chemical compounds implied in those processes change. We can also add the possible presence of clouds in the low atmosphere of Titan due to methane condensation that could be similar to our water clouds on Earth. The presence on Titan of liquid reservoirs such as lakes or seas of methane and ethane has also been suggested.

Meanwhile, those similarities between Titan and the Earth should not mask the major differences. For example, while the surface pressures are approximately the same, the mean ground temperatures are very different: -180°C on Titan compared to 20°C on Earth. This leads to a five times higher density on Titan than on the Earth at the ground level. Another difference appears if we compare the compositions of both atmospheres: nitrogen is the major compound for both of them but methane (CH_4) is second in Titan whereas nowadays on

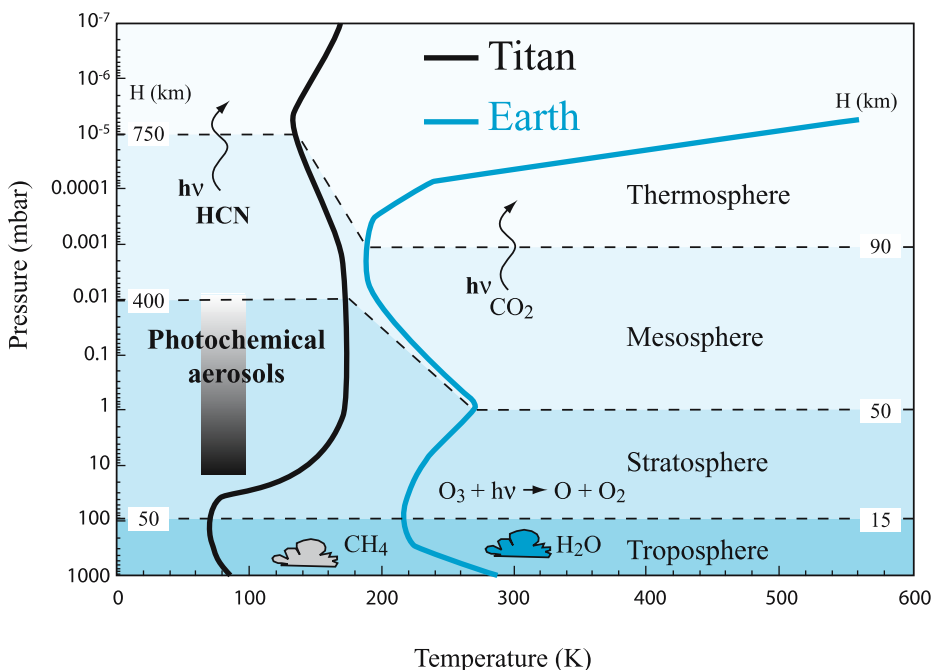


Fig. 13.18. Comparison between Earth's and Titan's atmospheric profiles

Earth it is molecular oxygen and water. Furthermore, carbon appears mostly on Earth in its oxidized state CO_2 . Water certainly played an important role in the appearance of life. Traces of water vapor have also been detected in the high atmosphere of Titan by Coustenis et al. (1998), using the ISO satellite (Infrared Space Observatory), but at the temperature of Titan's surface, this water is certainly solid. Nevertheless, Thompson and Sagan (1991) showed that meteoritic impacts on Titan could have led to ejecta containing water liquefied under the shock. According to this model, large quantities of liquid water could have persisted on the surface for a total duration of several thousands years. However, would this be enough for an aqueous organic chemistry to develop and produce prebiotic molecules? Finally, the presence of an underground ocean, as for example in Europa, opens other exobiological perspectives (Fortes 2000), but as already mentioned the question of the transport between this ocean and the surface remains a major unsolved problem.

In Titan's atmosphere, a solid phase is also present as dust particles. By condensation at their surface volatiles can be carried to the surface and possibly into a liquid methane/ethane surface. Therefore, Titan contains all the elements of what we can call a "geofluid" (Raulin 1997). This underlines the importance of the study of this satellite for exobiology. Indeed, this is the planetary body that is the closest to what could have been the Earth before the appearance

of life. Of course, the major difference with the primitive Earth is certainly the very low temperature of Titan, which implies the absence of liquid water. Nevertheless, the similitude between what could have been the primitive Earth and Titan, is one of the main reasons that motivated the launch of a probe called Huygens, which entered Titan's atmosphere in January 2005 to do in situ measurements.

Finally, one will note that even if in situ measurements bring data of fundamental importance on the composition, structure and dynamic of the atmosphere, they are done on a single trajectory of the space-time domain. Therefore, they have to be completed by teledetection measurements that allow to study phenomena at large scales with their time variations. This is done by the instruments onboard the Cassini spacecraft that has been put into orbit around Saturn in July 2004 for a minimum period of four years.

13.3.4 Cassini–Huygens Mission

The Cassini–Huygens space mission is a joint mission from ESA and NASA (Fig. 13.21). It is composed of an orbiter Cassini (NASA) and a descent probe Huygens (ESA). The whole set of instruments onboard is given in Table 13.9.

The main goal of the Huygens probe was to study the atmosphere and the surface of Titan. It has measured (in situ) the physical characteristics and chemical composition of the atmosphere and the surface. It was equipped with six main instruments. After being released by the orbiter Cassini, it entered Titan's atmosphere on January 14, 2005 and went through a first deceleration protected by its thermal shield. Its first small parachute opened around 180km and caused the opening of a second much larger parachute (Lebreton and Matson 1997). As soon as the main parachute was inflated the shield dropped out around the altitude of 160km. Forty-two seconds later, the first measurements started. Because of the limited capacity of the batteries (1.8kWh), a quarter later when the probe was around 125km altitude, the main parachute was replaced by a smaller one. With a maximum duration of 153 minutes predicted for the whole mission, the probe was supposed to do measurements on the surface for at least 3 minutes. Nevertheless, the reality exceeded all the predictions. After a 152-minute descent, the probe landed on Titan smoothly (4.5m/s on humid sand like material) and continued to emit toward Cassini for more than an hour (the probe actually continued to emit for several hours but no relay could send back the signal to Earth).

One of the key objectives of the Huygens probe was to determine the chemical species present under gaseous, solid or liquid phase in the atmosphere and on the surface. However, further than simply obtaining quantities the main goal was to study the coupling between those different phases in order to try to understand the evolution that occurred on Titan since its formation. It is this approach that should allow to gain new elements to understand the prebiotic chemistry that could have occurred on Earth before the emergence of life. In

Table 13.9. Ongoing missions

Mission	Space agency	Dates	Description and objectives	Instruments
Cassini Mass: 2523 kg Nominal power: 640 W	NASA/ ESA	10/15/1997: Launch 07/01/2004: Orbit insertion around Saturn	The first goal of the Cassini mission (NASA) is to transfer the descent probe Huygens (ESA) to Titan. Then it will stay in orbit around Saturn for a initial period of 4 years for a detail study of the rings, of the satellite surfaces including Titan, Saturn and Titan atmospheres including the spatial and temporal variability in terms of dynamic and chemical composition. Huygens probe has the specific mission to determine the physical and chemical characteristics of Titan atmosphere and surface	On board Cassini: Imaging system (ISS), magnetospheric imagery (MIMI), energy and charge of particles measurements (CAPS), magnetic field measurements (MAG), radio measurements (RPWS), cosmic dust analyzer (CDA), neutral and ionic mass spectrometer (INMS), radar infrared spectrometer (CIRS), ultraviolet spectrometer (UVIS), visible/IR spectroimager (VIMS)
Huygens Mass: 319 kg Nominal power: 250 W		01/14/2005: Huygens descent in Titan atmosphere (2H ₃₀)		On board Huygens: Atmospheric density, pressure and temperature profiles measurements (HASI), winds measurements (DWE), surface imaging and solar energy deposition in the atmosphere (DISR), Gas phase Chromatography coupled to mass spectrometry (GC-MS), collect et analysis of aerosols (ACP) and surface parameters measurements (SSP)

First results are being acquired and can be found on the site: <http://saturn.jpl.nasa.gov/home/index.cfm>

fact, if the “prebiotic chemist” is able to synthesize in the laboratory almost all biomacromolecules starting from hydrocarbons and nitriles observed on Titan, the question about what could happen on a planetary scale over geological time is still open. The questions that remain are (Raulin 1997):

- What complexity could achieve organic synthesis in a reduced atmosphere where liquid water is absent?
- What could have the importance of multiphasic processes?
- What is the influence of the physical conditions (temperature, energetic deposition and dynamics) on the chemistry?

The study of Titan (environment on which liquid water is absent, temperature very low, and energy sources diverse) can supply elements that answer those questions. The whole set of experiments on board the Huygens probe will contribute: the gas phase chromatograph coupled to a mass spectrometer (GCMS, Gas Chromatograph Mass Spectrometer), the collect and analysis of aerosols (ACP, Aerosol Collector Pyrolyser), the imager spectroradiometer (DISR, Descent Imager/Spectral Radiometer), the analysis of the surface (SSP, Surface Science Package) and the determination of the atmospheric structure (HASI, Huygens Atmospheric Structure Instrument).

a) GCMS

The instrument is composed of a chromatographic system made of three columns coupled to a mass spectrometer: one column was devoted to the measurement of permanent gases, one to light hydrocarbons (up to three carbons) and finally one is for heavier compounds up to eight nitrogen or carbons atoms (Niemann et al. 1997). The atmosphere has been analyzed during the entire descent by direct injection in the mass spectrometer (2–141 daltons). During the first half hour, a system was able to concentrate noble gases before their injection into the mass spectrometer. The chromatographic system itself was used on five different gaseous samples coming from five different altitudes. The first one was taken just after the start of the instruments around 150km. Two samples were then taken above 60km altitude where the concentration of complex compounds was expected to be the highest. The next analysis was made on a sample obtained by the pyrolysis of the aerosols collected by ACP (see below). At the minimum temperature of the tropopause, the analysis was devoted to the separation of N₂ and CO. Finally, since the descent time was nominal a last sample has been injected before impact. Moreover, since by chance the injector was correctly positioned, measurements of the surface composition have been done. The sensibility of the instrument was supposed to be very high of about ten parts per billion for most compounds. All the instrumentation worked correctly except the ion source attached to the separation of N₂ and CO. The first analysis of the measurements has shown an increase of the methane concentration occurring just after impact.

b) ACP

Two regions of the atmosphere were sounded by this instrument: the first one between 135 and 32km and the second one between 22 and 17km (Israel et al. 1997). The aerosols were collected on a filter, which was placed a few millimeters outside the probe, and then taken back inside the pyrolysis oven. The oven allowed a warming of the samples up to 600°C. The following sequence was planned: (1) collect during the first hour of the descent then warming at ambient temperature, 250°C and 600°C, then transfer to GCMS; and (2) collect a second sample above possible clouds then warm at the three temperatures and transfer to the GCMS. The main goal of those measurements was to determine the composition of the condensation nuclei by comparing the measurements obtained in Titan atmosphere with analogues produced in laboratory simulations. The relative composition of the condensates present on the aerosol surface should be obtained. Finally, the possible abundance of adsorbed volatile like CO in the aerosols might be determined. The system seems to have worked nominally.

c) DISR

During the whole descent, optical instrumentation of DISR (0.35–1.7 μm domain) measured the upwelling and downwelling solar flux in order to determine the net flux (Tomasko et al. 1997). The difference between net fluxes taken at two different altitudes allows to determine the energy deposited. This enables the understanding of the thermal balance of the atmosphere. Measuring the diffusion properties, for two colors and two polarizations, together with the extinction of the aerosols leads to the determination of the refractive index. Since this quantity is characteristic of a given material, it should be possible to constrain the chemical composition of Titan's aerosols. This optical instrumentation should allow to obtain the methane mixing ratio as a function of altitude by measuring its absorption and should also help to constraint the nature of the surface by determining its reflectivity. The measurements continued even in the last minutes of the descent when the probe was getting closer to the surface (where the atmosphere became more and more opaque) thanks to a 20W lamp attached to the instrument. DISR also includes two cameras that obtained images of the surface with a resolution going from a few hundred meters (like the orbiter) at the beginning of the descent to a few tenths of a centimeter when approaching the surface. The first analysis of those images showed the presence of geological structures characteristic of precipitation, erosion and fluvial activity similar to what is observed on Earth (Fig. 13.19). The surface is relatively flat and on the impact site, the presence of ice blocks of a few tens of centimeters can be observed (Fig. 13.20). The surface seems to be dry around the impact site but methane rainfall might have occurred recently.

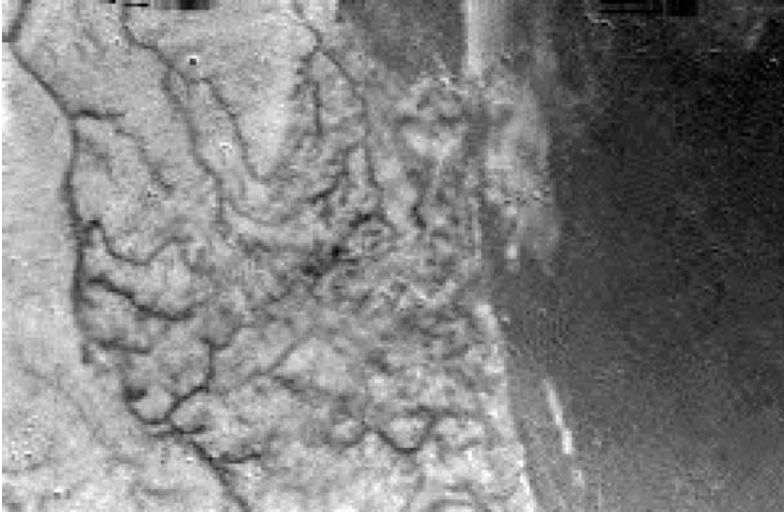


Fig. 13.19. This image was obtained by DISR at 16.2km above the Titan surface. The resolution is 40 meters per pixel. The presence of a fluvial network and a coastline can be observed (courtesy of ESA/NASA/JPL/University of Arizona)



Fig. 13.20. This image was obtained after the Huygens probe landed on the Titan surface. The surface is relatively flat and several ice blocks can be identified. The typical size of those blocks is fifteen centimeters (courtesy of ESA/NASA/JPL/University of Arizona)

d) HASI

The first goal of this instrumentation was to determine the density, pressure and temperature profiles all along the descent (Fulchignoni et al. 1997). However, this experiment also included the possibility to measure the electric field, the electric conductivity and even the sound waves through a microphone. Thus, it should be possible to determine the electronic density in the atmosphere but also deduce the presence of lightning storms if those are present around the descent trajectory. The conductivity of the surface could also be measured after landing which can help to constrain the nature of the surface. The first analysis showed that the measured pressure and temperature vertical profiles are very consistent with the one obtained by Voyager and no lightning was observed.

e) SSP

This experiment was composed of nine different instruments (Zarnecki et al. 1997). The set of instruments has been made to give information on the surface whether solid or liquid. If the surface had been liquid, the thermal conductivity, the density, the sound speed, the refracting index, and the permittivity could have been determined. Nevertheless, it was the accelerometer and the penetrometer that gave the interesting information since the surface was solid: granularity, penetration resistance and cohesion could be determined. An acoustic sounding device will give an idea of the topography of the surface. The sound speed, the temperature or the permittivity of the atmosphere were also determined during the descent

The instruments of the orbiter will also contribute. In fact, they allow a global view, which is essential to interpret the probe data. The two major instruments that supply data concerning Titan's atmospheric composition are:

f) CIRS

This instrument is an infrared Fourier transform spectrometer, which acquires data in the range of 10 to 1400cm^{-1} (1mm to $7\mu\text{m}$) (Flasar et al. 2004). This improved version of the IRIS spectrometer that was on the Voyager spacecrafts (extension of the wavelength domain, of the resolution and of the sensitivity) allows a mapping in three dimensions not only the chemical composition of Titan's atmosphere but also its temperature.

g) UVIS

This ultraviolet spectrometer is composed of two channels that study the radiations that are coming from Titan between 56 and 190nm (Esposito et al. 2003). It also obtains limb data from solar or stellar occultation. This should permit the determination of the vertical profile of numerous species in the high atmosphere where a large amount of the energy is deposited.

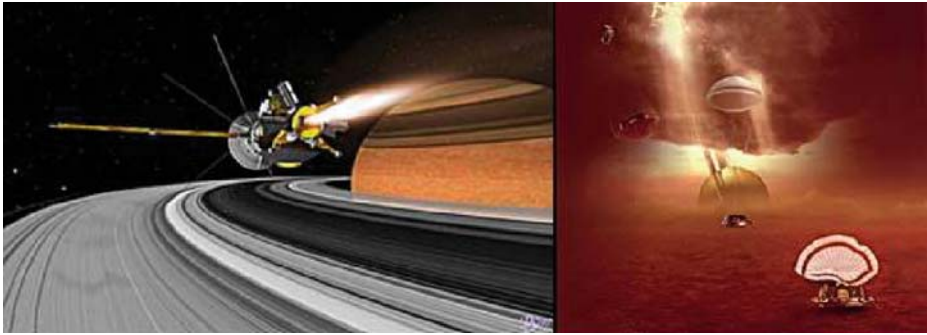


Fig. 13.21. Cassini–Huygens spacecraft arriving in the Saturn system and Huygens probe descent inside Titan’s atmosphere

The results of this space mission will certainly lead to a large breakthrough in our understanding of organic processes that could lead to the emergence of life on a planet having a reductive atmosphere and liquid water. However, even if the Cassini–Huygens mission will bring a very large amount of data it will certainly not answer all the questions concerning the complexity of Titan. Especially, since the descent was over one trajectory, it will not be possible to document the great complexity that already appears either on surface images or as a north-south asymmetry. In the future, it is possible and desirable that a new mission could be sent but this time with a mobile laboratory that could study different regions of Titan surface and different altitudes in the atmosphere (Lorentz 2001).

13.4 Mars Exploration

13.4.1 Mars Before Space Missions

Planet Mars, known since the prehistoric times (Egyptians were already talking about it in 4000 BC) has always attracted human beings. This is certainly partly due to its special trajectory in the sky, its aspect (its apparent diameter can vary by a factor greater than five), and its reddish color. On November 28, 1659, Huygens observed for the first time structures on the disk of Mars that changed with time (Fig. 13.22). From his observations, he deduced a rotational period of about 24 hours as on Earth. This gave rise to several phantasms that tended to assimilate the Martian world to the terrestrial one.

The observations made in the following centuries amplified those phantasms. Seas and “continents” appeared at the surface of the red planet. At the end of the nineteenth century, Lowell imagined that a channel network irrigates the surface of Mars to supply its inhabitants. It was necessary to wait for the beginning of space exploration to definitively drop the idea that a life similar to the one on Earth could be present.

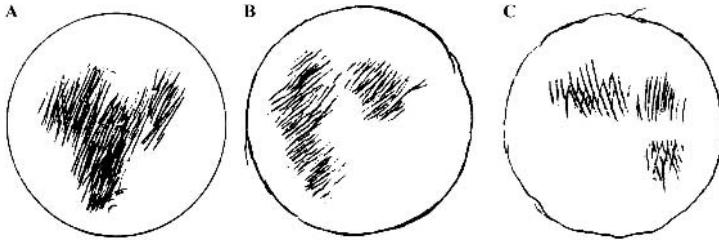


Fig. 13.22. Three drawings made by Huygens in 1659. The first one was drawn on November 28 at 7 pm (A), the second was done the same day but at 9:30 pm. (B). The last one was made three days later December 1 at 6:30 pm (C)

Another hot topic was the one of the nature of the polar caps and the presence of water. Polar caps that were observed for the first time by Cassini have been followed through the centuries and the observer had noted their appearance and disappearance with seasons. They were first supposed to be made of water by analogy with the Earth but also because the measured temperature was not low enough for carbon dioxide to be present in a condensed phase (Hess 1948). According to this theory, water should have been observed in the atmosphere. However, it was necessary to wait for the observations of Spinrad et al. (1963) to detect trace amount of water vapor, about 10 precipitable microns. Finally, the first space missions were needed to have a better idea of the water cycle on Mars.

The nature of the Martian atmosphere itself (pressure, composition, temperature) has been the subject of many discussions. In October 1947, Kuiper observed Mars in the wavelength region around $1.6\mu\text{m}$. Comparing this spectra with the one obtained from the moon that contains telluric carbon dioxide absorption lines, he showed that the absorption was stronger in Martian spectra (Kuiper 1952). Nevertheless, he could not deduce from his data the abundance of CO_2 because the bands were saturated. It was necessary to wait for the very high-resolution spectra obtained by Kaplan et al. (1964), in the 870nm domain where the bands are optically thin, to measure the abundance of CO_2 at the surface: 4mbar. The total surface pressure was still unknown at that time. Nevertheless, using observations around $2\mu\text{m}$, Kuiper (1963) and Sinton (1963) deduced that the total pressure was between 10 and 40 mbar. Therefore, for the first time, carbon dioxide appears as the main constituent of the Martian atmosphere.

13.4.2 The Beginning of Martian Exploration

Only eleven of the thirty or so missions sent to Mars effectively reached their goals. The lost missions have been reported in Table 13.12. As can be seen there are various reasons of failure, explosion at ignition, communication lost during Mars transfer or, more recently, communication lost during landing.

If the Soviets were the first to send a spacecraft to Mars, it is the US Mariner 4, launched November 28, 1964, which was the first to reach Mars on July 14, 1965. Thanks to its onboard camera, it took the first pictures of the Martian surface with a resolution of 4km. The ground appears riddled with craters, which lead to compare it with the moon, and seemed to indicate that the surface was as old as the planet itself (Chapman et al. 1969). But, only 1% of the surface had been pictured. The greatest breakthrough was done thanks to the radio occultation experiment that allowed for the first time to determine the surface pressure and the temperature. The pressure was of about 5 mbar (Kliore et al. 1965). Those measurements confirmed that carbon dioxide is the main element of a thin atmosphere. The next mission burst to Mars with Mariner 6 and 7 in 1969, followed by Mariner 9 in 1971 and finally the Viking probes in the mid 1970s, allowed to reach a more and more precise idea of the Martian environment using heavier and heavier equipment including more and more advanced technology (see Table 13.10).

Mariner 6 and 7 probes were equipped not only with a camera able to obtain images of 10% of the surface, but also an infrared spectrometer, an infrared radiometer and an ultraviolet spectrometer. The latest identified atomic hydrogen, atomic oxygen and carbon monoxide in the high atmosphere (Thomas 1971). The infrared measurements allowed to identify carbonic ice as the major component of the south pole (Herr and Pimentale 1969). The radiometer measured a temperature at the south pole of 148K, which led to confirm that it was composed of carbonic ice (Neugebauer et al. 1971). In fact, vapor pressure of CO₂ at this temperature is about 6 mbar, which is compatible with the measurements of Kliore et al. (1971). In the same conditions, the vapor pressure of water is only 10×10^8 mbar. Finally, thanks to the UV spectrometer, ozone was also discovered.

The Mariner 9 interplanetary probe stayed in orbit for a period of almost one terrestrial year. This allowed the cartography of almost all the Martian surface, the study of the atmospheric physicochemistry and its temporal variations over a period ranging from one day to one month. Imagery experiment showed an unsuspected diversity of the surface with the presence of volcanoes among which the highest, Olympus Mons, reaches 26km and canyons deeper than 6km like Valles Marineris. Valleys, which could have been formed by an erosion process, were also observed. Furthermore, water vapor was detected for the first time in Martian orbit by the infrared Fourier transform spectrometer IRIS, (Hanel et al. 1972) during the north summer. The condensation of this water vapor in the form of crystals was also observed (Curan et al. 1973). All those results pushed forward for an in situ mission able to bring into light the possible traces of biological activity on Mars.

The Viking project included two vehicles, Viking 1 and Viking 2, each one composed of an orbiter and a probe able to land on the Martian surface after a site selection. The main objectives of the Viking missions were the study of the Martian surface with high-resolution imagery, the characterization of the

Table 13.10. Past missions

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
MARINER 4 Mass: 260.7 kg	NASA	11/28/1964: Launch	First space probe arrived at Mars	Magnetometer, ion detector,	22 pictures were sent to the Earth. Surface pressure was estimated
Nominal power: 170 W		07/14/1965: Mars flyby	Take surface images; obtain data on micrometeoritic impacts, magnetic field and solar wind	radiometer, meteoritic particles detector and camera	between 4.1 and 7 mbar. No magnetic field observed
MARINER 6 Mass: 411.8 kg	NASA	02/25/1969: Launch	Study Mars surface and atmosphere to prepare future missions	Wild field and long focal camera, infrared spectrometer and radiometer,	M6: 75 pictures M7: 125 pictures
Nominal power: 449 W				ultraviolet spectrometer. The south pole is covered with carbonic ice	Surface pressure is estimated between 6 and 7 mbar
MARINER 7 Mass: 411.8 kg	NASA	03/27/1969: Launch			
Nominal power: 449 W					
MARS 2 Mass: 2265 kg	USSR	05/19/1971: Launch	First orbiter and lander mission. Study Mars surface and atmosphere to determine composition, temperature and topography.	Wild field and long focal camera, infrared spectrometer and radiometer, ultraviolet spectrometer,	Mars 2 lander was lost during atmospheric entry. Communication was lost with Mars 3 lander
MARS 3 Mass: 2265 kg	USSR	05/28/1971: Launch	Serve as relay to the Earth for the landers	surface modules	after 20s; 60 pictures were obtained

Table 13.10. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
MARINER 9 Mass: 997.9kg Nominal power: 449 W	NASA	05/30/1971: Launch	First orbiter mission around Mars. To obtain surface map, study the atmosphere together with its spatial, diurnal and secular variations	Camera, Infrared Fourier transform spectrometer. Ultraviolet spectrometer	Surface pressure is between 5 and 6 mbar. Atomic hydrogen and oxygen are detected in the high atmosphere More than 7000 images. Determination of the wind field and the atmospheric dynamic. Determination of the water content of the atmosphere. Detection of water ice clouds
MARS 5 Mass: 2270 kg	USSR	07/25/1973: Launch	Composition, structure and properties of Mars surface and atmosphere	Cameras, radiometer, visible and radio polarimeters, 5 photometers (study H at, Ly α ozone around, 250 nm visible albedo, H ₂ O and CO ₂ in the infrared)	After a few days breakdown of the transmission system. Nevertheless, a high water concentration (100 μ m) is measured close to Tharsis and an ozone layer is observed at high altitudes

Table 13.10. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
MARS 6 Mass: 635 kg	USSR	08/05/1973: Launch	First in situ probe. Study in situ Martian surface and atmosphere	Telephotometer, temperature and pressure sensors, wind measurements, accelerometer and mass spectrometer	Lost contact just before landing. Nevertheless, determination of the atmospheric profile below 25 km
VIKING 1 and 2 Mass: 997.9 kg Nominal power: 449 W	NASA	08/20/1975: Launch V1 09/09/1975: Launch V2	Orbiters: thermal and visible map of Martian surface. Study the water abundance in the atmosphere and its spatial, diurnal and secular variations. Landers: study the physical and chemical atmospheric parameters. Take pictures of the ground, obtain meteorological measurement of the surface, study the mineral and chemical, especially organic, composition of the surface, and search for the presence of life on Mars	Orbiters: camera, infrared spectrometer center on the water bands (MAWD), radio-meter. Landers: cameras, pressure and temperature sensors and anemometer, gas phase chromatography couple to a mass spectrometer. X fluorescence spectrometer, biological activity measurement experiments	Orbiters: more than 50 000 images obtained covering 87% of the Martian surface with a 200-m resolution. Determination of the water content in the atmosphere and its annual variations. Detection of a residual north polar cap made of water

Table 13.10. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
PHOBOS 2 Mass: 2600 kg	USSR	07/12/1988: Launch	Characterization of the plasma environment around Mars. Study Mars atmosphere and the surface, study Phobos surface	Among others: plasma analyzer, particles detector, infrared spectrometer, mass spectrometer, etc.	Landers: in situ determination of the temperature and pressure profiles during the entry phase. Meteorological measurements during 3 Martian years. Measure of the chemical and isotopic composition of the atmosphere. Determination of the inorganic composition of the surface with the presence of ferrosilicates. No organic matter detected Allowed to obtain the first mineralogical maps of Mars thanks to an infrared imaging spectrometer

Table 13.10. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
MARS PATHFINDER	NASA	12/04/1996: Launch	First rover on Mars. Technological mission	Station: imagers and meteorological station Rover: X spectrometer and cameras	16 500 pictures obtained by the station and 550 by the rover
Mass of the Rover: 10 kg					

structure and the composition of the atmosphere and the surface and the search for possible forms of life.

The orbiters mapped the entire surface of Mars with a 200-meter resolution. In addition to the recurrent dust storms, they observed the variations of the surface pressure all along the Martian year. This evidenced the presence of a CO₂ cycle, related to the decrease and the growth of the polar caps. The infrared spectrometer allowed to determine the water vapor content in the Martian atmosphere (MAWD, Mars Atmospheric Water Detectors) and to follow the atmospheric humidity global variations during the seasons due to the seasonal cycle of the water transport between the two poles. Finally, the coupling with the radiometer (IRTM, Infrared Thermal Mappers) results permitted the demonstration that if the south pole was mainly composed of CO₂ ices, the residual north polar cap visible during the summer was mainly composed of water ice.

The Viking 1 lander made continuous measurements in the region of Chryse Planitia on the Martian surface during six terrestrial years whereas the Viking 2 lander operated during two Martian years in Utopia Planitia (Fig. 13.23). Each probe was carrying not only a set of experiment to detect possible life forms on the Martian surface but also instruments to study the chemical, magnetic, apparent and physical properties of the surface and the atmosphere, and to achieve meteorological and seismic observations.

A small meteorological station supported by a one-meter mast was composed of three thermal sensors distributed along the height, a pressure gauge and an anemometer allowing the measure of both wind speed and direction. Those mea-



Fig. 13.23. One of the Viking landers with its mechanical arm deployed (photo courtesy of NASA/JPL)

surements acquired during more than three Martian years constituted a unique database for the study of Mars climate.

Inorganic chemical analysis of the surface were made by X fluorescence (XRFS). The obtained results were disconcerting. If Mars seemed to be constituted mainly of silicon and oxygen like the Earth, iron instead of aluminum was the next most abundant element.

Each of the Viking landers was also equipped with a gas chromatograph coupled with a mass spectrometer (CPG-SM). They were designed to determine the molecular and isotopic atmospheric composition and to try to detect organic matter in the ground. The atmosphere could also be directly injected in the mass spectrometer. This led to the determination of the isotopic abundance of five constituents. Solid samples were first pyrolysed at 500°C and then the gaseous residues transferred to the chromatograph. No organic compounds were identified down to the threshold of one part per billion (Biemann et al. 1977). This result was surprising since even if no endogen sources were present, the exogenous sources should have enriched the Martian soil in organic compounds. One explanation is that organic matter might have been destroyed by the combined action of radiation sources and oxidation.

The exobiological experiment, onboard the Viking landers (Fig. 13.23), was composed of three distinct experiments (Klein et al. 1972). The first one was the “pyrolytic release”. This experiment planned to detect carbonic gas assimilation (CO₂) either by photosynthetic mechanisms or in the dark. Martian samples were injected into a mixture of ¹²CO₂ and ¹⁴CO (with a ratio of 95:5). During a five day period, the sample could be illuminated by a Xenon lamp simulating the solar flux but filtered below 320nm to suppress UV. After this period, the residual atmosphere was evacuated and the sample pyrolysed in an oven at 625°C. The organic matter was then extracted and the amount of ¹⁴C determined. The procedure could be reproduced with water or in the dark. The results were difficult to interpret. First, experiments could not be reproduced: if the first experiment gave a positive result, the following one was negative. Furthermore, even if the sample was heated to high temperatures (175°C) the results would stay positive. The best explanation is that the positive experiment could result from a chemical compound (for example, the formation of a polymer like carbon suboxide (C₃O₂)), or from a physical process like the adsorption of CO₂ in the ground (Horowitz et al. 1977). The second experiment was the “gas Exchange”. This experiment was made to detect different gases that living organisms could reject after ingestion of nutriments. A nutriment solution was added to surface samples in contact with the Martian atmosphere. For the calibration, part of the sample was heated for sterilization. The determination of the gas composition present over the sample was done through GC-MS analysis. Here again, results were astonishing as they showed a high oxygen degassing and a low release of CO₂ and nitrogen. The release of oxygen was very quick and a second introduction of nutriment did not liberate more oxygen. Furthermore, the degassing occurred even in the dark or with the sterilized sample. Those re-

sults could be explained if one supposes that Martian soil contains very reactive oxidant molecules like superoxides (Oyama and Berdahl 1977). Finally, the third experiment was the “labeled release”. This experiment is similar to the previous one except that the carbon atoms of each compound inside the nutriment were radioactive. One detector could identify the outgassing radioactive CO_2 . A high outgassing of CO_2 was observed after the injection of nutriment. This release was not observed with the sterilized samples. So, one could conclude the presence of biological activity from this experiment only. Nevertheless, this last result can also be explained by non-biological processes. For example, the action of hydrogen peroxide (H_2O_2) could explain the release of carbon dioxide from the oxidation of the organic matter present in the nutriment (Huguenin et al. 1979).

The results of those experiments built to detect biological activity are still today subject to controversy because one can hardly deny the positive results of the label release experiment (Levin et al. 1981). Nevertheless, if microorganisms were really present inside the samples, organic matter should have been detected by the GC-MS experiment (Klein and Harold 1978). Consequently, most scientists agree to say that life is absent from the surface of Mars. Nevertheless, the question, if there is life on Mars, is still an open issue since some niches could exist in the subsurface.

13.4.3 Current Space Missions

Space missions currently in activity around Mars are resumed in Table 13.11. After the success of the 1970s and notably the Viking probe experiences, a long inactivity of almost twenty years occurred for the Martian program. In 1988, the Russian Space Agency put into orbit around Mars its first probe: Phobos 2. However, one had to wait until the end of the 1990s to see NASA land on Mars its first rover: Pathfinder. The mission is a true technological achievement and opens the way for a new Martian program. The missions that have followed had as a common objective to map the Martian surface and atmosphere to try to answer the question: when has there been water on Mars, in what state and where is it today? In fact, one of the essential needs for life seems to be liquid water. Consequently, we hope that following the water on Mars would lead us to the sources of life.

Within this objective, Mars Global Surveyor satellite brought its piece of the puzzle. The infrared imaging spectrometer allowed mapping the minerals present on the surface of Mars. Chistensen et al. (2000) identified the presence of hematite concentrated in a unique region: Meridiani Planum (Fig. 13.24). The presence of still water in the past could have allowed the precipitation of this iron oxide (Chistensen et al. 2004).

This observation led to the selection of this site for the landing of one of the Mars Explorating Rovers mission, MER B alias Opportunity. The other rover (MER A alias Spirit) landed in the Gusev Crater (Fig. 13.25). MER B confirmed

Table 13.11. In progress missions

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
MARS GLOBAL SURVEYOR Mass: 1030.5 kg Nominal power: 667 W	NASA	11/07/1996: Launch	Take images of the surface, study thermal emission and measure the topography, gravitational and magnetic field	Camera (MOC), infrared thermal spectrometer (TES), laser altimeter (MOLA)	A complete map of the surface is achieved by MOC at a 1.5-m resolution. The mineralogical composition is determined using TES and hematite is discovered. With MOLA, Martian topography is determined with absolute precision of 30m. Still active to relay transmission from other missions like the rovers
MARS ODYSSEY Mass: 376 kg	NASA	04/07/2001: Launch	Measure radiations in Martian environment to study potential risk for human explorer. Determine mineralogical composition and study the elemental	Radiation detector (MARIE), camera imaging spectrometer (THEMIS), gamma ray spectrometer (GRS) and neutron	Map of the hydrogen content of the subsurface (> 1m). Seems to show the presence of water ice in high latitudes and the presence of hy-

Table 13.11. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
MARS EXPRESS Beagle 2 Mass: 666 kg Nominal power: 460 W	ESA	06/02/2003: Launch	composition and hydrogen content of the surface Global geological map at high resolution (10m), mineralogical map, atmospheric composition map, study of the subsurface	detectors and spectrometer (NS and HEND) Stereo imager, (HRSC) mineralogy imaging spectrometer (OMEGA), UV/IR spectrometer (SPICAM), IRTF spectrometer (PFS), radar (MARSIS), high energy atoms and plasma analyzer (ASPERA)	drated minerals at lower latitudes. Still active to relay transmission from other missions like the rovers First results being released. See the site: http://www.esa.int/SPECIALS/ Mars-Express/index.html
MER A (Spirit) MER B (Opportunity)	NASA	06/10/2003: Launch MER A 07/07/2003: Launch MER B	Study Mars surface to search for past water traces	Camera, infrared spectrometer (mini-TES) Mossbauer, and X spectrometer (APX), microscope	MER A: landed in Gusev Crater, has covered 3.5 km in 9 months. MER B: landed in Terra Meridiani, travelled only 1.5 km, but confirmed the presence of hematite

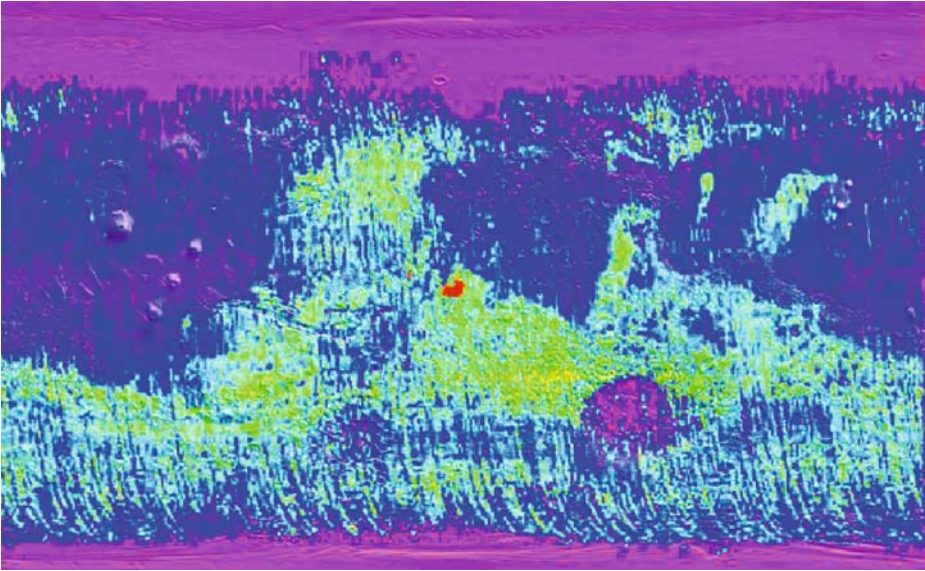


Fig. 13.24. Hematite map on Mars obtained by TES of the MGS mission (Christensen et al. 2000). Regions with high hematite concentration appear in red

the presence of hematite concentrated in small black spherules called “blueberries” (Moore 2004). The rover has also found jarosite, an iron sulfate. It could form if an atmosphere rich in carbon dioxide and sulfur dioxide is in contact with a large water mass rich in iron, (Fairén et al. 2004). Such conditions would preclude the formation of carbonates, which could explain their non-detection until now at the Martian surface. The presence of sulfates in Meridiani Planum has been confirmed by the Omega experiment onboard Mars Express mission (Gendrin et al. 2004). Furthermore, this experiment has allowed the identification of those minerals in and around Valles Marineris. The same instrument also put in light the presence of water ice at the south pole (Bibring et al. 2004).

The mapping of hydrogen present in the ground at less than one meter below the surface, either in the form of ice or in the form of hydrated minerals, has been obtained by the neutron detector and the gamma spectrometer of the Mars Odyssey mission (Fig. 13.25). A large concentration of ice appears at the poles and a low concentration at the equator except in two regions: Arabia Terra and Apollineris. In those regions, the water could be in the form of ice present several tenth of centimeter below a dehydrated layer.

Another instrument onboard Mars Express Mission has made a discovery of large exobiological implications: the infrared Fourier transform spectrometer observed methane (Formisano et al. 2004). This detection has also been confirmed by terrestrial observations (Krasnopolski et al. 2004; Mumma et al. 2004); this

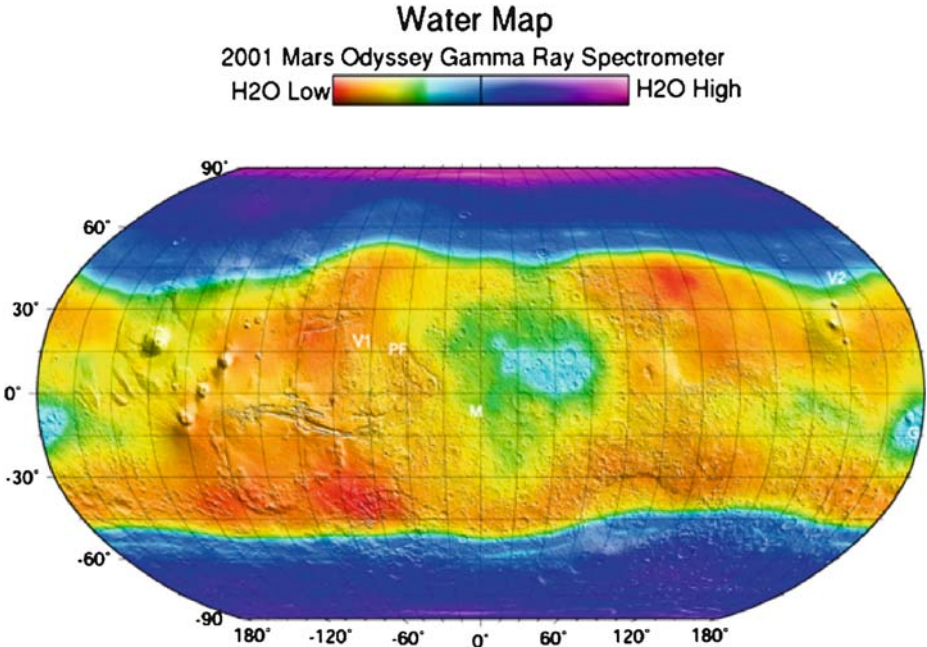


Fig. 13.25. Map created using the hydrogen measurements of the gamma spectrometer of the Mars Odyssey mission (Mitrofanov et al. 2002). If this hydrogen is really in the form of water, regions in *blue/violet* show the presence of more than 50% ice by volume in the ground. The significantly drier equatorial regions show two sites more hydrated around Arabia Terra (*at center*) and Apollineris (*on the side*). The localizations of the lander missions are indicated: Viking 1 (*VL1*), Viking 2 (*VL2*), Pathfinder (*PF*), Spirit at Gusev (*G*), and Opportunity at Meridiani (*M*)

methane has a very low abundance of one part per billion of carbon dioxide. Nevertheless, since the lifetime of methane in Mars atmosphere is of the order of one century, it has to be replenished from a source. On Earth, this source is mainly of biological origin, but Mars volcanism could be the explanation. In the near future, cartography of the methane content in the Martian atmosphere should provide some clues.

13.4.4 Future Exobiological Missions

If the results obtained by the exobiological experiment of Viking missions did not allow to definitely conclude on the presence of biological activity on Mars, they had the merit to exist in order to allow the criticism. This helps today to plan mission scenarios optimized for the search of life traces on Mars.

The first action has begun, since as we saw in the previous paragraph, the present program consists of trying to identify the places and the times where liquid water could exist on Mars. A second step will be to search for organic

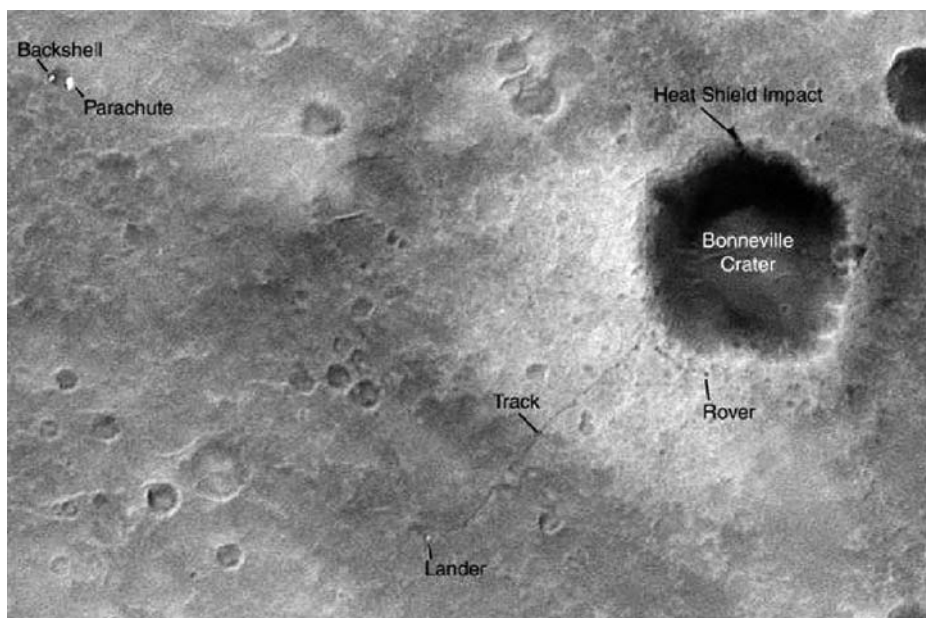


Fig. 13.26. Image acquired on March 30, 2004 by MOC of MGS (resolution 1.5×0.5 m/pixel). One can see the landing site of Spirit together with the track of the rover in direction of the Bonneville Crater during 85-sol (image courtesy of NASA/JPL/Malin Space Science Systems)

matter but such a target could be achieved only by in situ measurements. Future missions will also have to go further. They will have to look not only at the Martian surface but also to go deeper inside to explore the subsurface. The reason is that oxidation processes by molecules like hydrogen peroxide (H_2O_2) recently observed in the atmosphere (Clancy et al. 2004), or superoxides certainly lead to the destruction of organic matter on the surface (Yen et al. 2000). The thickness of the oxidized regolith has been evaluated theoretically but the uncertainties are such that this thickness can vary from one meter to several tenths of a meter (Zent 1998). Consequently, future missions will have to be able to drill at least one meter or even more.

The analytical system will have to be able to extract organic matter from the rock and analyze it to determine its composition. The pyrolysis/GC-MS technique could be improved compared to the one used on the Viking mission. First, derivatization could allow the analysis of refracting molecules that could appear in the degradation of the initial organic material. Then, the use of chiral chromatographic columns could allow the determination of the origin of molecules, either biotic or abiotic (Cabane et al. 2004). Isotopic measurements could also complete such analysis of the organic matter because isotopic ratio can help to identify biotic process. For example, on Earth carbon fractionation ($^{12}\text{C}/^{13}\text{C}$) in

Table 13.12. Lost missions

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
MARSNIK 1 Mass: 650 kg	USSR	10/10/1960: Launch	First known attempt of an interplanetary probe. Study interplanetary space between Mars and Earth. Study Mars and take surface images. Test long distance radio transmissions	Magnetometer, ions detector, radiometer, meteoritic particles detector, camera and spectroradiometer to study CH bands possibly indicative of life	Lost during launch, the stage did not ignite
MARSNIK 2 Mass: 650 kg	USSR	10/14/1960: Launch	Second attempt of an interplanetary probe. Same objectives as MARSNIK 1	See Marsnik 1	Lost during launch
SPUTNIK 22 Mass: 893.5 kg	USSR	10/24/1962: Launch	See Mars 1	See Mars 1	Explosion in Earth orbit
MARS 1 Mass: 893.5 kg	USSR	11/01/1962: Launch	Take surface images, obtain data on micrometeoritic impact and the magnetic field, determine the atmospheric structure and the possible presence of organic compounds	Magnetometer, ions detector, radiometer, meteoritic particles detector, camera and spectroradiometer to study ozone band	Communications lost during transit to Mars
SPUTNIK 24 Mass: 890 kg	USSR	11/04/1962: Launch	First attempt of a lander	Unknown	Explosion during Mars transit

Table 13.12. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
MARINER 3 Mass: 260.8 kg	NASA	11/01/1964: Launch	Take surface images, obtain data on micrometeoritic impact and the magnetic field, and the solar wind	Magnetometer, ions detector, radiometer, meteoritic particles detector, and camera	Communications lost during transit to Mars
ZOND 2 Mass: 893.5 kg	USSR	11/30/1964: Launch	See Mars 1	See Mars 1	Communications lost during transit to Mars
ZOND 3 Mass: 960 kg	USSR	07/18/1965: Launch	See Mars 1	See Mars 1	Launch with a trajectory not compatible with Mars flyby
MARS 1969A Mass: 4850 kg	USSR	03/27/1969: Launch	First attempt of a Mars orbiter	Camera, radiometer, water vapor detector, infrared, ultraviolet	Explosion during ignition of the third stage of the Proton launcher.
MARS 1969B Mass: 4850 kg	USSR	04/02/1969: Launch		and gamma spectrometer, mass spectrometer	Problem with the ignition of one of the launcher motors lead to explosion
MARINER 8 Mass: 558.8 kg	NASA	05/09/1971: Launch	Mars orbiter to measure temperature, pressure fields and the composition of the atmosphere and surface	Ultraviolet and infrared spectrometer and radiometer and camera	Defect of ignition of the last launcher stage

Table 13.12. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
COSMOS 419 Mass: 4650 kg	USSR	05/10/1971: Launch	Mars orbiter	See Mars 2	Defect of ignition of the last launcher stage
MARS 4 Mass: 2270 kg	USSR	07/21/1973: Launch	See Mars 5	See Mars 5	Defect of ignition of the retrorocket during Mars orbit insertion
MARS 7 Mass: 1200 kg	USSR	08/09/1973: Launch	See Mars 6	See Mars 6	The too early drop of the module led to miss Mars surface
PHOBOS 1 Mass: 2600 kg	USSR	07/07/1988: Launch	See Phobos 2	See Phobos 2	Communication lost during Mars transfer
MARS OBSERVER Mass: 1018 kg	NASA	09/25/1992	Determine elemental and mineralogical composition of the surface, topographic and gravitational and magnetic field measurements, study of volatiles an dust together with their spatial and seasonal variations	Camera, gamma spectrometer, laser altimeter, radio-meter, thermal infrared spectrometer, magnetometer	Communication lost during Mars transfer
MARS 96 Mass: 3159 kg	RSA	11/16/1996	See Mars Express	See Mars Express	Re-entry in Earth atmosphere for unknown reason

Table 13.12. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
NOZOMI Mass: 258 kg	ISAS	07/04/1998	Study the high atmosphere and its interactions with the solar wind	Camera, neutral and ionic mass spectrometer, particles counter, VUV imaging spectrometer, wave plasma detector and low frequency analyzer	In December 1998, Nozomi had problems with its propulsion system. The trajectory was then changed to save the mission. The trip will take four more years. But, the probe will flyby Mars December 14, 2003 without being able to be inserted into orbit
MARS CLIMATE ORBITER Mass: 338 kg	NASA	12/11/1998	Everyday climate study including surface changes due to atmospheric events, determination of temperature profiles, follow the water vapor and dust atmospheric content	Radiometer and cameras	Navigation error leading to the destruction of the probe during its orbit insertion on Mars

Table 13.12. (continued)

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
MARS POLAR LANDER Mass: 290 kg	NASA	01/03/1999	Record meteorological conditions of the south pole, analyze the surface deposit, drill trench to search for seasonal layers, determine surface composition	Descent imager, laser system (LIDAR) to study atmospheric aerosols, meteorological station, camera, spectrometer to study volatiles after pyrolysis	Communications never established after landing
DEEP SPACE 2 Mass: 3.6 kg	NASA	01/09/1999	Impactor allowing to test the presence of water ice under the surface, determine the properties of the sub surface. Measure temperature and pressure	Accelerometer, water detector, conductivity measurement	Communications never established after impact
BEAGLE 2 Mass: 33 kg	ESA	06/02/2003	Geological, mineralogical, geochemical and climatological studies, search for biosignatures	Drilling system, gas chromatography and mass spectrometry, microscope, camera, X fluorescence spectrometer and Mossbauer	Communications never established after landing

Table 13.13. Future Martian missions

Mission	Space agency	Dates	Description and objectives	Instruments	Main results
MARS RECONNAISSANCE ORBITER	NASA	08/12/2005: Launch	Characterize surface, sub-surface and atmosphere, identify sites for future in situ missions	High resolution imaging, two wild field cameras, IR-visible imaging spectrometer, radio-meter and radar	It is adjusting its orbit using aerobraking to reach low-altitude, near-circular orbits at a mean altitude of 280 km above the Martian surface
PHOENIX	NASA	Launch 2007	In situ studies of the north pole, geological history of water, search for potential habitable zone at the interface ground/ice	Descent imager, camera, robotized arm, mass spectrometer, microscopes, chemical and meteorological measurements	Low cost scout mission (heritage from Mars Polar Lander) being prepared
MARS SCIENCE LABORATORY	NASA	Launch in 2009	Mobile laboratory on Mars surface	Call for proposal being released	Call for proposal being released

favor of ^{12}C is of the order of 20 to 30% in the biomass compared to inorganic carbon (Brack et al. 1999). Such measurement set will probably be present on the future MSL mission (Mars Science Laboratory) planned in 2009 (see Table 13.13).

Nevertheless, with all the in situ detection effort of life traces, the sign of possible Martian life could be so subtle that the equipment necessary for its detection would be much too heavy and complex to be put on a space probe. The solution would then be to bring samples collected on Mars back to the Earth to study them in the laboratory. This type of mission initially planned for the end of the decade is actually not expected before 2015.

13.5 Conclusion

To date, there is nothing that could allow us to figure out if we will ever know whether the Earth is the only inhabited planet in the universe. However, astrobiology has gone a long way since the very first experiments by Miller in 1950 and since then, questions related to astrobiology are feeding the exploration of our solar system, and even beyond with ambitious programs for the observation of extrasolar planets (TPF, Darwin). The answer is maybe already at hand, by a rock analyzed with a Martian rover, or further away, under a few kilometers of ice in the Jupiter neighborhood. . . This chapter deals with the preliminary steps of the astrobiological exploration of the solar system, we are currently witnessing. We focused on the search of organic matter and liquid water, but none of them is a clear indicator of past or present life. We did not discuss the next issue about measurement(s) (chemical, geological, spectroscopic), which could allow us to claim that we have finally detected life on another planet. The scientific debate about it is still at its very first stages, and opens the most fascinating prospects.

Acknowledgements

The authors wish to thank Antoine Jolly for a careful proofreading of the manuscript; his many comments and suggestions were a great help to improve the text. We also thank the editors of the book: Muriel Gargaud, Philippe Clays and Hervé Martin for their comments, encouragements and corrections. Hervé Martin did a wonderful job in processing the figures.

References

- Altwegg, K., H. Balsiger, and J. Geiss, Composition of the volatile material in Halley's coma from in situ measurements, *Space Science Reviews*, 90, 3–18, 1999.
- Atreya, S.K., T.M. Donahue, and W.R. Kuhn, Evolution of a nitrogen atmosphere on Titan, *Science*, 201, 611–613, 1978.

- Barbier, B., A. Chabin, D. Chaput, and A. Brack, Photochemical processing of amino acids in Earth orbit, *Planetary and Space Science*, 46 (4), 391–398, 1998.
- Barbier, B., O. Henin, F. Boillot, A. Chabin, D. Chaput, and A. Brack, Exposure of amino acids and derivatives in the Earth orbit, *Planetary and Space Science*, 50, 353–359, 2002.
- Bernstein, M.P., S.A. Sandford, L.J. Allamandola, S. Chang, and M.A. Scharberg, Organic Compounds Produced By Photolysis of Realistic Interstellar and Cometary Ice Analogs Containing Methanol, *The Astrophysical Journal*, 454, 327–344, 1995.
- Bibring, J.P., Y. Langevin, F. Poulet, A. Gendrin, B. Gondet, M. Berthé, A. Soufflot, P. Drossart, M. Combes, G. Bellucci, V. Moroz, N. Mangold, B. Schmitt, and the OMEGA team, Perennial water ice identified in the south polar cap of Mars, *Nature*, 428, 6983, 627–630, 2004.
- Biemann, K., J. Oro, P. Toulmin, L.E. Orgel, A.O. Nier, D.M. Anderson, D. Flory, A.V. Diaz, D.R. Rushneck, and P.G. Simmonds, The search for organic substances and inorganic volatile compounds in the surface of Mars, *Journal of Geophysical Research*, 82, 4641–4658, 1977.
- Bland, P.A., T.B. Smith, A.J.T. Jull, F.J. Berry, A.W.R. Bevan, S. Cloudt, and C.T. Pillinger, The flux of meteorites to the Earth over the last 50 000 years, *Monthly Notices of the Royal Astronomical Society*, 283, 551, 1996.
- Blank, J.G., G.H. Miller, M.J. Ahrens, and R.E. Winans, Experimental shock chemistry of aqueous amino acid solutions and the cometary delivery of prebiotic compounds, *Origins of Life and Evolution of the Biosphere*, 31, 15–51, 2001.
- Bockelée-Morvan, D., D. Gautier, F. Hersant, J.-M. Huré, and F. Robert, Turbulent radial mixing in the solar nebula as the source of crystalline silicates in comets, *Astronomy and Astrophysics*, 384, 1107–1118, 2002.
- Bockelée-Morvan, D., J. Crovisier, M.J. Mumma, and H.A. Weaver, The composition of cometary volatiles, In: *Comets II*, edited by M. Festou, H.U. Keller, and H.A. Weaver, University of Arizona Press, Tucson, AZ, 2004.
- Boillot, F., A. Chabin, C. Buré, M. Venet, A. Belsky, M. Bertrand-Urbaniak, A. Delmas, A. Brack, and B. Barbier, The Perseus exobiology mission on MIR: behavior of amino acids and peptides in Earth orbit, *Origins of Life and Evolution of the Biosphere*, 32, 359–385, 2002.
- Botta, O., and J.L. Bada, Extraterrestrial organic compounds in meteorites, *Surveys in Geophysics*, 23, 411–467, 2002.
- Brack, A., La chimie de l'origine de la vie, In: *Les traces du vivant*, edited by M. Gargaud, D. Despois, J.-P. Parisot, and J. Reisse, pp. 61–81, Presses Universitaires de Bordeaux, Bordeaux, 2003.
- Broadfoot, A.L., B.R. Sandel, D.E. Shemansky, J.B. Holberg, G.R. Smith, D.F. Strobel, J.C. McConnell, S. Kumar, D.M. Hunten, S.K. Atreya, T.M. Donahue, H.W. Moos, J.L. Bertaux, J.E. Blamont, R.B. Pumphrey and S. Linick, *Science*, 212, 206–211, 1981.
- Cabane, M., P. Coll, C. Szopa, G. Israël, F. Raulin, R. Sternberg, P. Mahaffy, A. Person, C. Rodier, R. Navarro-González, H. Niemann, D. Harpold, W. Brinckerhoff, Did life exist on Mars? Search for organic and inorganic signatures, one of the goals for “SAM” (sample analysis at Mars), *Advances in Space Research*, 33, 12, 2240–2245, 2004.
- Caldwell, J.J., Ultraviolet observations of small bodies in the solar system by OAO-2, *Icarus*, 25, 384–396, 1975.

- Carr, M.H., M.J.S. Belton, C.R. Chapman, M.E. Davies, P. Geissler, R. Greenberg, A.S. McEwen, B.R. Tufts, R. Greeley, and R. Sullivan, Evidence for a subsurface ocean on Europa, *Nature*, 391, 363, 1998.
- Christensen, P.R. et al., Detection of crystalline hematite mineralization on Mars by the Thermal Emission Spectrometer, *Journal of Geophysical Research*, 105, 9632–9642, 2000.
- Christensen, P.R., and S.W. Ruff, Formation of the hematite-bearing unit in Meridiani Planum: Evidence for deposition in standing water. *Journal of Geophysical Research*, 109, E8, 2004.
- Chyba, C.F., P.J. Thomas, L. Brookshaw, and C. Sagan, Cometary delivery of organic molecules to the early earth, *Science*, 249 (July), 249–373, 1990.
- Clancy, R.T., B.J. Sandor, and G.H. Moriarty-Schieven, A measurement of the 362 GHz absorption line of Mars atmospheric H₂O₂. *Icarus*, 168, 1, 116–121, 2004.
- Conklin, E.K., B.L. Ulich and J.R. Dickel, 3-mm Observations of Titan, *Bulletin American Astronomical Society*, 9, 471, 1977.
- Corliss, J.B., J.A. Baross, and S.E. Hoffman, An hypothesis concerning the relationship between submarine hot spring and the origin of life on Earth, *Oceanologica Acta*, N° SP, Proceedings of the 26th Geological Congress, 59–69, 1981.
- Cottin, H., C. Szopa, and M.H. Moore, Production of hexamethylenetetramine in photolyzed and irradiated interstellar cometary ice analogs, *The Astrophysical Journal Letters*, 561 (1), L139–L142, 2001.
- Cottin, H., Y. Bénilan, M.-C. Gazeau, and F. Raulin, Origin of cometary extended sources from degradation of refractory organics on grains: polyoxymethylene as formaldehyde parent molecule, *Icarus*, 167, 397–416, 2004.
- Coustonis, A., A. Salama, E. Lellouch, T. Encrenaz, G.L. Bjoraker, R.E. Samuelson, T. de Graauw, H. Feuchtgruber, M.F. Kessler, *Astronomy and Astrophysics*, 336, L85–L89, 1998.
- Cronin, J.R., and S. Pizzarello, Enantiomeric excesses in meteoritic amino acids, *Science*, 275 (14 February), 951–955, 1997.
- Curran, R.J., B.J. Conrath, R.A. Hanel, V.G. Kunde, and J.C. Pearl, Mars: Mariner 9 spectroscopic evidence for H₂O ice clouds, *Science* 175, 381–383, 1973.
- Danielson, R.E., J.J. Caldwell and D.R. Larach, An inversion in the atmosphere of Titan, *Icarus*, 20, 437–443, 1973.
- Despois, D., and H. Cottin, Comets: potential sources of prebiotic molecules for the early Earth, In: *Lectures in Astrobiology I*, chap. 8, edited by M. Gargaud, B. Barbier, H. Martin, and J. Reisse, Springer, Berlin Heidelberg New York, 2005.
- Despois, D., Les comètes, sources potentielles de molécules pour la terre primitive et les planètes, In: *L'environnement de la Terre Primitive*, edited by M. Gargaud, D. Despois, and J.-P. Parisot, pp. 53–77, Presses Universitaires de Bordeaux, Bordeaux, 2001.
- Eberhardt, P., and D. Krankowsky, The electron temperature in the inner coma of comet P/Halley, *Astronomy and Astrophysics*, 295, 795, 1995.
- Ehrenfreund, P., and S.B. Charnley, Organic molecules in the interstellar medium, comets and meteorites: a voyage from dark clouds to the early earth, *Annual Review of Astronomy and Astrophysics*, 38, 427–483, 2000.
- Encrenaz, T., B. Bézard, T.K. Greathouse, M.J. Richter, J.H. Lacy, S.K. Atreya, A S. Wong, S. Lebonnois, F. Lefèvre, and F. Forget, Hydrogen peroxide on Mars: evidence for spatial and temporal variations, *Icarus* 170, 424–429, 2004.

- Esposito, L.W., et al., The Cassini ultraviolet imaging spectrograph investigation, *Space Science Reviews*, 115 (1–4), 299–361, 2004.
- Fairén, A.G., D. Fernández-Remolar, J.M. Dohm, V.R. Baker, and R. Amils, Inhibition of carbonate synthesis in acidic oceans on early Mars, *Nature*, 431, 7007, 423–426, 2004.
- Flasar, M., et al., Exploring the Saturn system in the thermal infrared: the composite infrared spectrometer, *Space Science Reviews*, 115 (1–4), 169–297, 2004.
- Formisano, V., S. Atreya, T. Encrenaz, N. Ignatiev, and M. Giuranna, Detection of methane in the martian atmosphere, *Science*, 306, 1756–1761, 2004.
- Forterre, P., A la recherche des formes de vie terrestre les plus “primitives”: impasses et progrès, In: *L’environnement de la Terre Primitive*, edited by M. Gargaud, D. Despois, and J.-P. Parisot, pp. 399–416, Presses Universitaires de Bordeaux, Bordeaux, 2001.
- Fortes, A.D., Exobiological implications of a possible ammonia-water ocean inside Titan, *Icarus*, 146, 444–452, 2000.
- Fulchignoni, M., F. Angrilli, G. Bianchini, A. Bar-Nun, M.A. Barucci, W. Borucki, M. Coradini, A. Coustenis, F. Ferri, R.J. Grard, M. Hamelin, A.M. Harri, G.W. Leppelmeier, J.J. Lopez-Moreno, J.A.M. McDonnell, C. McKay, F.M. Neubauer, A. Pederson, G. Picardi, V. Pironello, R. Pirjola, R. Rodrigo, C. Sshwingenschuh, A. Seiff, H. Svedhem, E. Thrane, V. Vanzani, G. Visconti and J.C. Zarnecki., Huygens: Science, payload and Mission, ESA SP-1177, 163–195, 1997.
- Geiss, J., K. Altwegg, H. Balsiger, and S. Graf, Rare atoms, molecules and radicals in the coma of P/Halley, *Space Science Reviews*, 90, 253–268, 1999.
- Gendrin, A., J.P. Bibring, B. Gondet, Y. Langevin, N. Mangold, J.F. Mustard, F. Poulet, and C. Quantin. Identification of sulfate deposits on Mars by Omega/Mars Express. Proceedings of the 2nd Conference on Early Mars, 11–15 October 2004, Jackson Hole, WY.
- Gillett, F.C., Further observations of the 8–13 micron spectrum of Titan, *The Astrophysical Journal*, 201, L41–L43, 1975.
- Gillett, F.C., W.J. Forrest and K.M. Merrill, 8–13 Micron Observations of Titan, *The Astrophysical Journal*, 184, L93–L95, 1973.
- Greenberg, J.M., What are comets made of? A model based on interstellar dust, In: *Comets*, edited by L.L. Wilkening, pp. 131–163, University of Arizona Press, Tucson, AZ, 1982.
- Hanel, R., B. Conrath, F.M. Flasar, V. Kunde, W. Maguire, J. Pearl, J. Pirraglia, R. Samuelson, L. Herath, M. Allison, D. Cruikshank, D. Gautier, P. Gierasch, L. Horn, R. Koppany, and C. Ponnampertuma, *Science*, 212, 192–200, 1981.
- Hanel, R., B. Conrath, W. Hovis, V. Kunde, P. Lowman, W. Maguire, J. Pearl, J. Pirraglia, C. Prabhakara, and B. Schlachman, Investigation of the Martian environment by infrared spectroscopy on Mariner 9, *Icarus* 17, 423–442, 1972.
- Hennet, R.J.C., N.G. Holm, and M.H. Engel, Abiotic synthesis of amino acids under hydrothermal conditions and the origin of life: a perpetual phenomenon, *Naturwissenschaften*, 79, 361–365, 1992.
- Herr, K.C., G.C. Pimental, Infrared absorptions near three microns recorded over the polar CAP of Mars, *Science*, 166, 496–499, 1969.

- Hersant, F., D. Gautier, and J.-M. Huré, A two-dimensional model for the primordial nebula constrained by D/H measurements in the solar system: implications for the formation of giant planets, *Astrophysical Journal*, 554, 391–407, 2001.
- Horowitz, N.H., G.L. Hobby, J.S. Hubbard, Viking on Mars: the carbon assimilation experiments, *Journal of Geophysical Research*, 82, 4659–4662, 1977.
- Huebner, W.F., D.C. Boice, and A. Korth, Halley's polymeric organic molecules, *Advances in Space Research*, 9 (2), 29–34, 1989.
- Huebner, W.F., First polymer in space identified in Comet Halley, *Science*, 237, 628–630, 1987.
- Huguenin, R.L., K.J. Miller, W.S. Harwood, Frost-weathering on Mars: experimental evidence for peroxide formation, *Journal of Molecular Evolution*, 14, 103–132, 1979.
- Hunten, D.M., *The Saturn System*, edited by D.M. Hunten and D. Morison, NASACP 2068, 113–126, 1978.
- Iro, N., D. Gautier, F. Hersant, D. Bockelée-Morvan, and J.I. Lunine, An interpretation of the nitrogen deficiency in comets, *Icarus*, 161, 511–532, 2003.
- Israel, G., H. Niemann, F. Raulin, W. Riedler, S. Atreya, S. Bauer, M. Cabane, E. Chassefière, A. Hauchecorne, T. Owen, C. Sablé, R. Samuelson, J.P. Torre, C. Vidal-Majar, J.F. Brun, D. Coscia, R. Ly, M. Tintignac, M. Steller, C. Gelas, E. Condé, and P. Millan, Huygens: science, payload and mission, ESA SP-1177, 59–84, 1997.
- Kasting, J.F., Earth's early atmosphere, *Science*, 259, 920–926, 1993.
- Khare, B.N. and C. Sagan, Red clouds in reducing atmospheres, *Icarus*, 20, 311, 1973.
- Kissel, J., and F.R. Krueger, The organic component in dust from Comet Halley as measured by the PUMA mass spectrometer on board Vega 1, *Nature*, 326 (April), 755–760, 1987.
- Kissel, J., F.R. Krueger, J. Silén, and B.C. Clark, The cometary and interstellar dust analyzer at comet 81P/Wild 2, *Science*, 304, 1774–1776, 2004.
- Kivelson, M.G., K.K. Khurana, C.T. Russell, M. Volwerk, R.J. Walker, and C. Zimmer, Galileo magnetometer measurements: a stronger case for a subsurface ocean at Europa, *Science*, 289, 1340–1343, 2000.
- Klein, H.P., The Viking biological experiments on Mars, *Icarus*, 34, 3, 666–674, 1978.
- Klein, H.P., J. Lederberg, A. Rich, Biological experiments: the Viking Mars Lander, *Icarus*, 16, 139, 1972.
- Kliore, A., D.L. Cain, G.S. Levy, V.R. Eshleman, G. Fjeldbo, and F.D. Drake, Occultation experiment: results of the first direct measurement of Mars's atmosphere and ionosphere, *Science*, 149, 1243–1248, 1965.
- Krasnopolsky, V.A., J.-P. Maillard, and T. Owen. Detection of methane in the martian atmosphere: evidence for life? *Icarus* 172, 537–547, 2004.
- Krueger, F.R., A. Korth, and J. Kissel, The organic matter of comet Halley as inferred by joint gas phase and solid phase analyses, *Space Science Reviews*, 56, 167–175, 1991.
- Krueger, F.R., and J. Kissel, The chemical composition of the dust of comet P/Halley as measured by PUMA on board Vega 1, *Naturwissenschaften*, 74, 312–316, 1987.
- Kunde, V.G., A.C. Ainkin, R.A. Hanel, D.E. Jennings, W.C. Maguire, and R.E. Samuelson, *Nature*, 292, 686–688, 1981.
- Lebreton, J.P., and D. Matson, Huygens: science, payload and mission, ESA SP-1177, 1997.

- Levin, G.V., P.A. Straat, A search for a nonbiological explanation of the Viking labeled release life detection experiment. *Icarus*, 45, 2, 494–516, 1981.
- Lindal, G.F., G.E. Wood, H.B. Hotz, D.N. Sweetnam, V.R. Eshleman and G.L. Tyler, *Icarus*, 53, 348–363, 1983.
- Lorenz, R.D., Exo-/astrobiology. Proceedings of the 1st European Workshop, ESA SP-496, 215–218, 2001.
- Low, F.J., Planetary radiation at infrared and millimeter wavelengths, *Lowell Observatory Bulletin*, 6(128), 184–187, 1965.
- Lunine, J.I., S. Engel, R. Bashir, and M. Horanyi, Sublimation and reformation of icy grains in the primitive solar nebula, *Icarus*, 94, 333–344, 1991.
- Lutz, B.L., C. de Bergh and T. Owen, Titan: discovery of carbon monoxide in its atmosphere, *Science*, 220, 1374–1375, 1983.
- Maguire, W.C., R.A. Hanel, D.E. Jennings, V.G. Kunde and R.E. Samuelson, *Nature*, 292, 683–686, 1981.
- Marten, A., D. Gautier, L. Tanguy, A. Lecacheux, C. Rosolen and G. Paubert, *Icarus*, 76, 558–562, 1988.
- Maurette, M., Carbonaceous micrometeorites and the origin of life, origins of life and evolution of the biosphere, *Origins of Life and Evolution of the Biosphere*, 28, 385–412, 1998.
- Maurette, M., La matière extraterrestre primitive et les mystère de nos origines, In: *L'environnement de la Terre Primitive*, edited by M. Gargaud, D. Despois, and J.-P. Parisot, pp. 99–127, Presses Universitaires de Bordeaux, Bordeaux, 2001.
- McCord, T.B., G.B. Hansen, and C.A. Hibbitts, Hydrated salt minerals on Ganymede's surface: evidence of an ocean below, *Science*, 292, 1523–1525, 2001.
- McCord, T.B., G.B. Hansen, D.L. Matson, T.V. Johnson, J.K. Crowley, F.P. Fanale, R.W. Carlson, W.D. Smythe, P.D. Martin, C.A. Hibbitts, J.C. Granahan, and A. Ocampo, Hydrated salt minerals on Europa's surface from the Galileo near-infrared mapping spectrometer (NIMS) investigation, *Journal of Geophysical Research*, 104, 11827–11852, 1999.
- Meier, R., P. Eberhardt, D. Krankowsky, and R.R. Hodges, The extended formaldehyde source in comet P/Halley, *Astronomy and Astrophysics*, 277, 677–691, 1993.
- Meierhenrich, U.J., G.M.M. Caro, J.H. Bredehöft, E.K. Jessberger, and W.H.-P. Thiemann, Identification of diamino acids in the Murchison meteorite, *Proceedings of the National Academy of Sciences*, 101, 9182–9186, 2004.
- Miller, S.L., The production of amino acids under possible primitive Earth conditions, *Science*, 117, 528–529, 1953.
- Mitchell, D.L., R.P. Lin, C.W. Carlson, A. Korth, H. Rème, and D.A. Mendis, The origin of complex organic ions in the coma of comet Halley, *Icarus*, 98, 125–133, 1992.
- Mitrofanov et al., Maps of subsurface hydrogen from the High Energy Neutron Detector, *Mars Odyssey*, *Science*, 297, 78–81, 2002.
- Möller, G., and W.M. Jackson, Laboratory studies of polyoxymethylene: application to comets, *Icarus*, 86, 189–197, 1990.
- Moore, M.H., and T. Tanabe, Mass spectra of sputtered polyoxymethylene: implications for comets, *The Astrophysical Journal*, 365, 1990.
- Moore, J.M., Mars blueberry fields for ever, *Nature*, 428, 6984, 711–712, 2004.

- Mumma, M.J., R.E. Novak, M.A. Di Santi, B.P. Bonev, N. Dello Russo, Detection and mapping of methane and water on Mars, *Bulletin American Astronomical Society*, 36, 1127–1127, 2004.
- Muñoz Caro, G.M., and W.A. Schutte, UV-photoprocessing of interstellar ice analogs: new infrared spectroscopic results, *Astronomy and Astrophysics*, 412, 121–132, 2003.
- Ness, F.N., M.H. Acuña, R.P. Lepping, J.E.P. Connerney, K.W. Behannon, L.F. Burlaga and F.M. Neubauer, Magnetic field studies by Voyager 1: preliminary results at Saturn, *Science*, 212, 211–216, 1981.
- Neugebauer, G., G. Miinch, H. Kieffer, S.C. Chase, and E. Miner, Mariner 1969 infrared radiometer results: temperatures and thermal properties of the Martian surface. *The Astronomical Journal*, 76, 719, 1971.
- Niemann, H. et al., Huygens: science, payload and mission, ESA SP-1177, 85–107, 1997.
- Oro, J., and C.B. Cosmovici, Comets and life on the primitive Earth, in *astronomical and biochemical origins and the search for life in the universe*, edited by C.B. Cosmovici, S. Bowyer, and D. Werthimer, pp. 97–120, *Proceedings of the 5th International Conference on Bioastronomy*, Bologna, Italy, 1997.
- Oyama, V.I., B.J. Berdahl, The Viking gas exchange experiment results from Chryse and Utopia surface samples, *Journal of Geophysical Research*, vol. 82, 4669–4676, 1977.
- Pierazzo, E., and C.F. Chyba, Cometary delivery of biogenic elements to Europa, *Icarus*, 157, 120–127, 2002.
- Pizzarello, S., and J.R. Cronin, Non-racemic amino acids in the Murray and Murchison meteorites, *Geochimica et Cosmochimica Acta*, 64, 329–338, 2000.
- Podolak, M., and D. Prialnik, 26-Al and liquid water environments in comets, In: *Comets and the Origin and Evolution of Life*, edited by P.J. Thomas, C.F. Chyba, and C.P. McKay, Springer, Berlin Heidelberg New York, 1997.
- Pollack, J.B., Greenhouse models of the atmosphere of Titan, *Icarus*, 19, 43–58, 1973.
- Prinn, R.G., and B.J. Fegley, Solar nebula chemistry: origin of planetary, satellite and cometary volatiles, In: *Origin and Evolution of Planetary and Satellite Atmospheres*, edited by S.K. Atreya, J.B. Pollack, and M. Matthews, pp. 78–136, University Of Arizona Press, Tuscon, AZ, 1989.
- Raulin, F., Huygens: science, payload and mission, ESA SP-1177, 219–229, 1997.
- Raulin, F., Chimie prébiotique: expériences de simulation en laboratoire et "vérité terrain", In: *L'environnement de la Terre Primitive*, edited by M. Gargaud, D. Despois, and J.-P. Parisot, pp. 343–360, Presses Universitaires de Bordeaux, Bordeaux, 2001.
- Reisse, J., and J. Cronin, Chiralité et origine de l'homochiralité, In: *Les traces du vivant*, edited by M. Gargaud, D. Despois, J.-P. Parisot, and J. Reisse, pp. 83–113, Presses Universitaires de Bordeaux, Bordeaux, 2003.
- Rodier, C., O. Vandenabeele-Trambouze, R. Sternberg, D. Coscia, P. Coll, C. Szopa, F. Raulin, C. Vidal-Madjar, M. Cabane, G. Israel, M.F. Grenier-Loustalot, M. Dobrijevic, and D. Despois, Detection of martian amino acids by chemical derivatization coupled to gas chromatography: in situ and laboratory analysis, *Advances in Space Research*, 27, 195–199, 2001.
- Rosenbauer, H., S.A. Fuselier, A. Ghielmetti, J.M. Greenberg, F. Goesmann, S. Ulamec, G. Israel, S. Livi, J.A. MacDermott, T. Matsuo, C.T. Pillinger, F. Raulin, R. Roll, and W. Thiemann, The Cosac experiment on the lander of the Rosetta mission, *Advances in Space Research*, 23, 333–340, 1999.

- Samuelson, R.E., W.C. Maguire, R.A. Hanel, V.G. Kunde, D. Jennings, Y.L. Yung and A.C. Aikin, CO₂ on Titan, *Journal of Geophysical Research*, 88, 8709–8715, 1983.
- Schenk, P.M., Thickness constraints on the icy shells of the galilean satellites from a comparison of crater shapes, *Nature*, 417, 419–421, 2002.
- Selsis, F., and J.-P. Parisot, L'atmosphère primitive de la Terre et son évolution, In: *L'environnement de la Terre Primitive*, edited by M. Gargaud, D. Despois, and J.-P. Parisot, pp. 217–233, Presses Universitaires de Bordeaux, Bordeaux, 2001.
- Smith, B.A., J. Boyce, G. Briggs, A. Bunker, S.A. Collins, C.J. Hansen, T.V. Johnson, J.L. Mitchell, R.J. Terrile, M. Carr, A.F. Cook II, J. Cuzzi, J.M. Pollack, G.E. Danielson, A. Ingersoll, M.E. Davies, G.E. Hunt, H. Masursky, E. Shoemaker, D. Morrison, T. Owen, C. Sagan, J. Veverka, R. Strom and V.E. Suomi, Encounter with Saturn: Voyager 1 imaging science results, *Science*, 212, 163–191, 1981.
- Smith, G.R., D.F. Strobel, A.L. Bradfoot, B.L. Sandel, D.E. Shemansky and J.B. Holberg, Titan's upper atmosphere: composition and temperature from the EUV solar occultation results, *Journal Geophysical Research*, 87, 1351–1359, 1982.
- Sotin, C., J.W. Head, and G. Tobie, Europa: tidal heating of upwelling thermal plumes and the origin of lenticulae and chaos melting, *Geophysical Research Letters*, 29, 74–1, 2002.
- Stoks, P.G., and A.W. Schwartz, Nitrogen-heterocyclic compounds in meteorites: significance and mechanisms of formation, *Geochimica et Cosmochimica Acta*, 45, 563–569, 1981.
- Stoks, P.G., and A.W. Schwartz, Uracil in carbonaceous meteorites, *Nature*, 282, 709–710, 1979.
- Stone, E.C. and E.D. Miner, Voyager 1 encounter with the Saturnian system, *Science*, 212, 159–163, 1981.
- Strobel, D.F. and D.E. Shemansky, EUV emission from Titan's upper atmosphere: Voyager 1 encounter, *Journal of Geophysical Research*, 87, 1361–1368, 1982.
- Strobel, D.F., The photochemistry of hydrocarbons in the atmosphere of Titan, *Icarus*, 21, 466–470, 1974.
- Szopa, C., R. Sternberg, F. Raulin, and H. Rosenbauer, What can we expect from the in situ chemical investigation of a cometary nucleus by gas chromatography: first results from laboratory studies, *Planetary and Space Science*, 51, 863–877, 2003.
- Thomas, G.E., Neutral composition of the upper atmosphere of Mars as determined from the Mariner UV spectrometer experiments, *Journal of Atmospheric Sciences*, 28, 6, 859–868, 1971.
- Thompson, W.R., and C. Sagan, Symposium on Titan, ESA SP, 167–176, 1991.
- Tomasko, M.G. et al., Huygens: science, payload and mission, ESA SP-1177, 109–138, 1997.
- Trafton, L.M., On the possible detection of H₂ in Titan's atmosphere, *The Astrophysical Journal*, 175, 285–293, 1972a.
- Trafton, L.M., The bulk composition of Titan's atmosphere, *The Astrophysical Journal*, 175, 295–306, 1972b.
- Tyler, G.L., V.R. Eshleman, J.D. Anderson, G.S. Levy, G.S. Lindal, G.E. Wood and T.A. Croft, Radio science investigations of the Saturn system with Voyager 1: preliminary results, *Science*, 212, 201–206, 1981.
- Veverka, J., Titan: polarimetric evidence for an optically thick atmosphere, *Icarus*, 18, 657–660, 1973.

- Whipple, F.L., A comet model. I. The acceleration of Comet Encke, *The Astrophysical Journal*, 111, 375–394, 1950.
- Wolman, Y., S.L. Miller, J. Ibanez, and J. Oro, *Science*, 174, 1039, 1971.
- Yanagawa, H., and K. Kobayashi, An experimental approach to chemical evolution in submarine hydrothermal systems, *Origins of Life and Evolution of the Biosphere*, 22, 147–159, 1992.
- Yen, A.S., S.S. Kim, M.H. Hecht, M.S. Frant, B. Murray, Evidence that the reactivity of the Martian soil is due to superoxide ions, *Science*, 289, 5486, 1909–1912, 2000.
- Zarnecki, J.C., M. Banaszkiwicz, M. Bannister, W.V. Boynton, P. Challenor, B. Clark, P.M. Daniell, J. Delderfield, M.A. English, Fulchignoni, M., J.R.C. Garry, J.E. Geake, S.F. Green, B. Hathi, S. Jaroslowski, M.R. Leese, R.D. Lorenz, J.A.M. McDonnell, N. Merrywether-Clarke, C.S. Mill, R.J. Miller, G. Newton, D.J. Parker, P. Rabetts, H. Svedhem, R.F. Turner and M.J. Wright, *Huygens: science, payload and mission*, ESA SP-1177, 177–195, 1997.
- Zellner, B., The polarization of Titan, *Icarus*, 18, 661–664, 1973.
- Zent, A.P., On the thickness of the oxidized layer of the Martian regolith. *Journal of Geophysical Research*, 103, E13, 31491–31498, 1998.
- Zimmer, C., K.K. Khurana, and M.G. Kivelson, Subsurface oceans on Europa and Callisto: constraints from Galileo magnetometer observations, *Icarus*, 147, 329–347, 2000.

14 Quantum Astrochemistry: Numerical Simulation as an Alternative to Experiments

Yves Ellinger and Françoise Pauzat

Theoretical studies of chemical processes taking place in extreme physical conditions are a clear need of the astrophysical community that is actually confronted with situations almost impossible to reproduce in laboratory experiments. However, these non-standard chemistries are at the very basis of our understanding of the evolution of our Universe, from the synthesis of primordial molecules to the formation of terrestrial planets, from the primary atmosphere to the emergence of life.

Many questions arise which cannot be answered by experimental means. It happens, for example, in the search for new molecules whose mechanisms of formation and destruction are unknown or in the characterization of meteorites whose accretion processes are not established. It is also true of the relations between the structures of the ices and their surface chemistry. Most of the answers can be obtained from theoretical studies, which are the only alternative in a number of problems. At the frontiers of astrophysics, chemistry and life sciences, exobiology is a field widely open to numerical simulations.

Exobiology (astrobiology for some) is by nature interdisciplinary. In this type of research, it is important to understand the ideas behind quantum calculations. To make the most out of quantum chemistry, it is also necessary to know the strengths and weaknesses of the theoretical methods together with the actual limitations in the possible applications. It is the purpose of this chapter to present the basic principles and principal methods of this discipline. It will be illustrated by the results of a number of theoretical studies realized at LET-MEX (Laboratory for Theoretical Study of Extreme Media) in the domain of interstellar chemistry.

14.1 The Methods of Quantum Chemistry

14.1.1 Definition of Quantum Chemistry Calculations and Approximations

The goal of quantum chemistry calculations is to determine the *behavior* of electrons and nuclei within a molecule. It is done by using mathematical functions capable of describing quantum phenomena from first principles and the universal constants of physics. Whatever the choices made, evaluation of wave func-

tions or evaluation of electronic densities, all methods are based on solving the Schrödinger equation. In this equation, all the particles and all the interactions between them are taken into account. The equation to solve is

$$\check{H}(r, R)\check{\Psi}(r, R) = \varepsilon\check{\Psi}(r, R) \quad (14.1)$$

where $\check{H}(r, R)$ is the complete Hamiltonian of the system depending on the electronic r and nuclear R coordinates. This operator is expressed as

$$\check{H}(r, R) = T_e + T_A + V_{eA} + V_{ee} + V_{AA} \quad (14.2)$$

where T_e and T_A are the kinetic energy operators for the electrons and nuclei, respectively, and

$$T_e = \sum_{i=1}^N -\frac{1}{2}\nabla_i^2 \quad (14.3)$$

$$T_A = \sum_{a=1}^A -\frac{1}{2}\nabla_a^2. \quad (14.4)$$

The terms V_{eA} , V_{ee} and V_{AA} are the operators describing the Coulombic attraction between electrons and nuclei, the repulsion between the electrons and the repulsion between the nuclei given by

$$V_{eA} = \sum_{i=1}^N \sum_{a=1}^A -\frac{Z_a}{|R_a - r_i|} \quad (14.5)$$

$$V_{ee} = \sum_{i=1}^N \sum_{j>i}^N \frac{1}{|r_j - r_i|} \quad (14.6)$$

$$V_{AA} = \sum_{a=1}^A \sum_{b>a}^A -\frac{Z_a Z_b}{|R_a - R_b|} \quad (14.7)$$

The wave function in the Schrödinger equation $\check{\Psi}(r, R)$ is a function depending on all the electronic and nuclear (space and spin) parameters and ε is the total internal energy of the system under consideration (namely the total electronic energy minus the translation energy).

Solving the Schrödinger equation is not a simple process since the energy to be calculated has contributions from various types of movements, motions of electrons, vibrations and rotations of the nuclear frame, all entangled in a subtle way. In a rigorous way, the Schrödinger equation can only be solved exactly

for a two-electron atom. In all other cases, it is necessary to make approximations. Generally, the quantities so neglected are re-introduced as perturbations, a posteriori (for example, magnetic interactions). The quantitative description of a polyatomic entity then requires several approximations. General textbooks referred to at the end of this volume may help the reader to get a better insight into the original developments forming the basis of quantum chemistry as it stands today.

14.1.1.1 Decoupling Electronic from Nuclear Motions

The first approximation made amounts to decoupling the motion of the electrons from that of the nuclei. It is known as the *Born–Oppenheimer approximation*. It relies on the fact that nuclei are much heavier than electrons with the consequence that electrons are able to adjust their positions, instantaneously, to those of the nuclei. In these conditions, electrons remain in quasistationary states during vibrations and rotations of the nuclei so that the dynamic equations can be partitioned into an electronic part and a rovibrational part. Such an approximation amounts to neglecting the coupling terms between the two types of motions. The total wave function of the system can then be rewritten in the form of a product of a function describing only the nuclear motions (rotations and vibrations) $\Theta(R)$ with an electronic function $\Psi(r, R)$

$$\check{\Psi}(r, R) = \Psi(r, R)\Theta(R) . \quad (14.8)$$

In this context, the Schrödinger equation reduces to an electronic equation in which the nuclear coordinates intervene as parameters

$$H(r, R)\Psi(r, R) = E(R)\Psi(r, R) \quad (14.9)$$

where $\Psi(r, R)$ and $E(R)$ are characteristic of the eigenelectronic state for the geometry considered. The motion of the nuclei, which is at the origin of the IR and rovibrational spectra is treated, a posteriori, by solving a Schrödinger equation in which the electronic energy acts as a potential.

14.1.1.2 Independence of Electrons

The second approximation considers electrons as *independent particles*, meaning that every electron moves in the average field created by the others. It leads to solving, for every electron, a one-electron Schrödinger equation

$$h_i\varphi_i = e_i\varphi_i . \quad (14.10)$$

The one-electron wave function φ_i that describes the motion of electron i is called the spin-orbital and only depends on the space and spin coordinates of a single electron. In this case, the wave function of the electronic system is

a single product of spin-orbitals whose associated energy is the sum of the orbital energies. The fact that electrons are fermions and indiscernible particles, led to writing the wave function as an antisymmetrized product of one-electron wave functions which satisfies the Pauli principle.¹ For an N -electron system, one has

$$\Phi = \frac{1}{\sqrt{N!}} \sum_p \varepsilon_p P_p \varphi_1(1) \varphi_2(2) \dots \varphi_N(N) \quad (14.11)$$

an expression better known as the Slater determinant or electronic configuration,

$$\Phi = \frac{1}{\sqrt{N!}} \|\varphi_1(1) \varphi_2(2) \dots \varphi_N(N)\|. \quad (14.12)$$

The exchange terms between electrons appear naturally when developing the energy over the permutations associated with this type of wave function.

14.1.1.3 Electronic Correlation

The third approximation can be seen as a correction to the previous one. Considering electrons as independent is obviously incorrect when looking at the formation or rupture of chemical bonds or describing excited states. The example of the homolytic scission of a C–H bond in methane (Fig. 14.1) according to the reaction $\text{CH}_4 \rightarrow \text{CH}_3 + \text{H}$ is illustrative.

The C–H bond wave function Φ_{CH} can be written, with explicit spin functions α and β of the electrons as

$$\Phi_{\text{CH}} = \frac{1}{\sqrt{2}} \|\varphi_{\text{CH}}(1) \varphi_{\text{CH}}(2)\| (\alpha\beta - \beta\alpha) \quad (14.13)$$

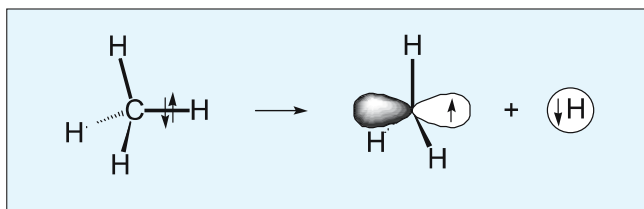


Fig. 14.1. Homolytic scission of a C–H bond in methane. Active electrons are represented by \uparrow and \downarrow for α and β spins, respectively

¹ Two particles (fermions) cannot exist in the same quantum state. In the present context, it means that two electrons with the same spin cannot occupy the same orbital. If the two electrons had the same spin, then the determinant would be equal to zero. The wave function would not exist! ε_p will be set to +1 or –1 according to the parity of the P_p permutations of the electrons.

where

$$\varphi_{\text{CH}} = \frac{1}{\sqrt{2}} (X_{\text{C}} + X_{\text{H}}) \quad (14.14)$$

is expressed in the form of a linear combination of atomic functions X_{C} and X_{H} that describe the electron on carbon and hydrogen. At the limit of dissociation

$$\begin{aligned} \Phi_{\text{CH}} &= \frac{1}{\sqrt{2}} [X_{\text{C}}(1)X_{\text{C}}(2) + X_{\text{C}}(1)X_{\text{H}}(2) + X_{\text{C}}(2)X_{\text{H}}(1) + X_{\text{H}}(1)X_{\text{H}}(2)] (\alpha\beta - \beta\alpha) \\ &\text{H}_3\text{C}^- + \text{H}^+ + \text{H}_3\text{C}^\bullet + \text{H}^\bullet + \text{H}_3\text{C}^+ + \text{H}^\bullet + \text{H}_3\text{C}^+ + \text{H}^- . \end{aligned} \quad (14.15)$$

This wave function contains several components of different characters, an ionic part with positive and negative fragments H_3C^- and H_3C^+ and a covalent part with radical H_3C^n that represent the physical solution to the problem.

A second example, illustrated on Fig. 14.2, is the fragmentation reaction of methylene CH_2 according to $\text{CH}_2 \rightarrow \text{C} + \text{H}_2$.

Another important property of wave functions is that they must belong to the same symmetry group as the system of nuclei that characterize the geometry of the molecule (C_{2v} in this example²). This is true for one-electron as well as for N -electron total wave functions.

In C_{2v} symmetry, every molecular orbital transforms in itself (1) or in its opposite (-1) when applying the different symmetry operations (Table 14.1).

The wave functions³ of the reactants and products are given by

$${}^3\text{CH}_2 : 1a_1^2 2a_1^2 1b_2^2 3a_1^1 1b_1^1 \quad \text{and} \quad ({}^3\text{P})\text{C} + ({}^1\Sigma_g)H_2 : 1a_1^2 2a_1^2 3a_1^2 4a_1^1 1b_1^1 .$$

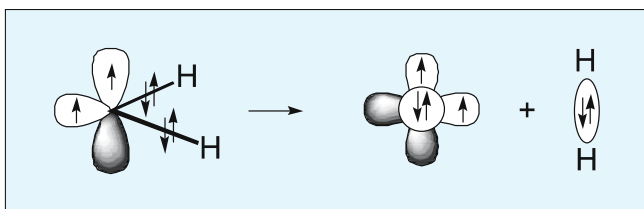


Fig. 14.2. Fragmentation of methylene leading to molecular hydrogen

² C_{2v} symmetry is defined by four operations. When applied to the nuclei, they reproduce the molecular skeleton identically. These symmetry operations are: identity (E), reflection with respect to HCH (σ_v), reflection with respect to the bisector plan of HCH (σ'_v), and rotation with respect to the axis bisecting HCH (C_2).

³ The notation $3a_1^2$ means that the third orbital of a_1 symmetry is occupied by two electrons (α and β); $4a_1^1$ means that the fourth orbital of a_1 symmetry is occupied by one electron (α).

Table 14.1. Symmetries of one-electron wave functions of reactants and products in the reaction $\text{CH}_2 \rightarrow \text{C} + \text{H}_2$

Operation	E	C_2	σ_v	σ'_v	Symmetry
Reactant: CH_2					
1sC	1	1	1	1	a_1
1pC	1	1	1	1	a_1
CH + CH	1	1	1	1	a_1
CH-CH	1	-1	1	-1	b_2
$2p_z\text{C}$	1	-1	-1	1	b_1
Products: $\text{C} + \text{H}_2$					
1sC	1	1	1	1	a_1
2sC	1	1	1	1	a_1
H-H	1	1	1	1	a_1
$2p_x\text{C}$	1	1	1	1	a_1
$2p_z\text{C}$	1	-1	-1	1	b_1

Both configurations belong to symmetry 3B_1 , although the electrons are occupying orbitals with different⁴ symmetries. Clearly, these two examples show that the single-configuration description (by a single determinant) is not satisfactory.

In fact, the motion of an electron does not depend only on the average electrostatic field within the molecule. It depends, explicitly, on the instantaneous positions of the other electrons. This effect, well-known under the name of electronic correlation, can be taken into account by the configuration interaction (CI) method that mixes the ground state configuration with appropriate excited configurations. A learning example is given by the treatment of electronic correlation in the helium atom whose ground state electronic configuration is

$$\Phi_1 = \frac{1}{\sqrt{2}} \|\varphi_s(1)\varphi_s(2)\|(\alpha\beta - \beta\alpha). \quad (14.16)$$

In the CI procedure, the total wave function Ψ is taken as a linear combination of Φ_1 with excited configurations in which two electrons are promoted in the $2p_x$, $2p_y$ and $2p_z$ orbitals

$$\Psi = \sum_{i=1}^4 c_i \Phi_i \quad (14.17)$$

⁴ The symmetry of the total wave function is given by the product of the symmetry representations Γ_i of the occupied spin-orbitals. For any orbital occupied by two electrons $\Gamma_i \cdot \Gamma_i = a_1$; in addition, $\Gamma_i \cdot a_1 = \Gamma_i$. In the present example of C_{2v} symmetry, one has: $b_1 \cdot b_2 = a_2$; $b_1 \cdot a_2 = b_2$; $b_2 \cdot a_2 = b_1$.

with

$$\Phi_2 = \frac{1}{\sqrt{2}} \|\varphi_{2p_x}(1)\varphi_{2p_x}(2)\|(\alpha\beta - \beta\alpha) \quad (14.18)$$

$$\Phi_3 = \frac{1}{\sqrt{2}} \|\varphi_{2p_y}(1)\varphi_{2p_y}(2)\|(\alpha\beta - \beta\alpha) \quad (14.19)$$

$$\Phi_4 = \frac{1}{\sqrt{2}} \|\varphi_{2p_z}(1)\varphi_{2p_z}(2)\|(\alpha\beta - \beta\alpha) \quad (14.20)$$

in which the φ_i s are exponential functions of exponents ζ , characteristic of helium

$$\varphi_s = e^{-\zeta r}; \quad \varphi_{2p_x} = xe^{-\zeta r}; \quad \varphi_{2p_y} = ye^{-\zeta r}; \quad \varphi_{2p_z} = ze^{-\zeta r}.$$

This CI development makes it possible to understand how electronic correlation is introduced in the calculations by means of doubly-excited configurations (Fig. 14.3).

Neglecting normalization, one may write $c_1 = 1$ and $c_2 = c_3 = c_4 = 2c$. Under these conditions

$$\Psi = e^{-\zeta(r_1+r_2)} [1 + 2c(r_1 \cdot r_2)] (\alpha\beta - \beta\alpha), \quad (14.21)$$

that is,

$$\Psi = e^{-\zeta(r_1+r_2)} [1 + c(r_1^2 + r_2^2 - r_{12}^2)] (\alpha\beta - \beta\alpha). \quad (14.22)$$

It can be seen that correlation between electrons 1 and 2 appears explicitly in the wave function via r_{12} , namely the instantaneous distance between the two electrons (Fig. 14.3).

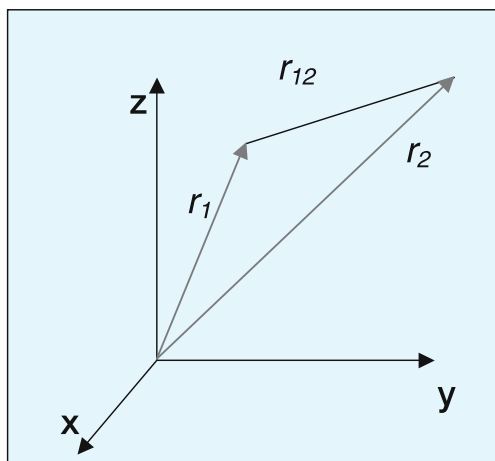


Fig. 14.3. Illustration of electronic correlation between two electrons defined by their respective positions r_1 and r_2 , with respect to the He nucleus

14.1.1.4 LCAO Approximation

The fourth approximation is directly linked to the practical aspect of molecular calculations. For pragmatic reasons, it is more convenient that φ_i spin-orbitals have analytic forms easily tractable for numerical calculations. One solution is to express these molecular orbitals as linear combinations of atomic orbitals X_p , the LCAO approximation

$$\varphi_i = \sum_p c_p X_p . \quad (14.23)$$

The idea behind the LCAO expansion is that, even in a molecular environment, the description of the isolated atoms remains a valid approximation.

The first atomic orbitals used in quantum chemistry calculations are known as Slater orbitals. They are expressed in the general form

$$f_{\text{Slater}} = x^n y^m z^p e^{-\zeta r} \quad (14.24)$$

where the exponent ζ is characteristic of the atom considered and r is the distance between the electron and the nucleus. These orbitals are derived from hydrogenoid functions with a correct asymptotic decrease when r goes to infinite distances. The main inconvenience of these exponential functions is the very large amount of calculation needed for the evaluation of the interaction energy terms between electrons. For that reason, a new atomic basis of Gaussian type has been developed, whose general form is

$$f_{\text{Gauss}} = x^n y^m z^p e^{-\alpha r^2} . \quad (14.25)$$

Several Gaussian functions are necessary to reach the quality of a single Slater function. The atomic orbital is now written in the form of a linear combination of elementary Gaussian functions

$$X = \sum_g c_g f_{\text{Gauss}} \quad (14.26)$$

whose exponents α and coefficients c_g (referred to as contraction coefficients) are determined in preliminary calculations of the ground state of the isolated atoms. These atomic calculations are generally done within the independent particle approximation. More recently, correlation effects have been taken into account in the atomic treatments to obtain better atomic basis sets, more appropriate to configuration interaction. In order to increase flexibility in the atomic representation, it is of common use to add extra functions with quantum number l greater than that of the functions occupied in the atomic ground state (polarization orbitals often marked by “*”) and diffuse functions (often marked by “+”). These types of basis sets are implemented in most quantum chemistry codes and widely used by the community.

14.1.2 Wave Function-Based Methods

In the methods based on the calculation of the wave function, the level of theory associated with a given treatment can be seen as its capability to numerically approach the exact wave function. It depends on a quadruple choice that can be expressed in the form of four equations interlocked like a nest of Russian dolls.

1. *Choice of the atomic basis set:*

$$X_p = \sum_g G_g \quad (14.27)$$

where G_g are generally Gaussian type functions centered on the atoms and previously determined in independent atomic calculations. The basis is formed of the orbitals occupied in the separate atoms (core and valence orbitals), to which are added polarization orbitals allowing the deformation of the electronic cloud under the action of the electric field created by the electronic distribution on the surrounding atoms. These are the functions referred to above with azimuthal quantum number greater than the last occupied valence shell (for example, $3d$ for C and $4f$ for Fe atoms).

The general idea behind this choice is that the basis set should be able to cover the entire space the wave function will spread over. In the same way, Rydberg orbitals are added for describing highly excited states and diffuse orbitals are added for negative ions, since electrons are much farther from the nuclear backbone in these systems.

2. *Choice of the spin-orbitals:*

$$\varphi_i = \sum_p c_{pi} X_p \quad (14.28)$$

where c_{pi} are the coefficients giving the participation of the various atoms in the molecular levels φ_i . They are calculated or are a priori chosen in order to represent the chemical bonds and lone pairs.

3. *Choice of electronic configurations:*

$$\Phi_K = 1/\sqrt{N!} \|\varphi_1(1)\varphi_2(2) \dots \varphi_N(N)\| \quad (14.29)$$

where φ_i are the one-electron wave functions used in the construction of Slater determinants or electronic configurations Φ_K , taking into account spatial symmetry and spin multiplicity.

4. *Choice of the basis of electronic configurations:*

$$\Psi = \sum_K d_K \Phi_K \quad (14.30)$$

where the total electronic wave function Ψ is expanded in the form of a linear combination of electronic configurations Φ_K built on the φ_i molecular orbitals previously calculated.

Once these choices have been made, the energy E and the associated wave function Ψ are calculated as

$$E = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle}. \quad (14.31)$$

Two different types of methods are employed, i.e., variational methods based on the minimization of the energy E , or perturbational methods whose principle is to improve an already satisfying zeroth-order wave function Ψ^0 .

14.1.2.1 Variational Methods

As a consequence of the four approximations presented above, the energy can be written as a function of two series of coefficients: the coefficients c_{pi} of the atomic orbitals and the coefficients d_K of the electronic configurations. In order to minimize the energy with respect to variations in the c_{pi} and d_K parameters, two conditions must be satisfied in the variational methods:

$$\frac{\partial E}{\partial c_p} = 0 \quad \text{and} \quad \frac{\partial E}{\partial d_K} = 0.$$

Following the choices that are made, one can use self-consistent field methods, CI methods, etc.

a) Self-Consistent Methods

By fixing the coefficients d_K of the basis of electronic configurations such that

$$\{d_K\} = \{1, 0, 0, \dots, 0\} \quad \text{and} \quad \frac{\partial E}{\partial c_p} = 0$$

one obtains the category of single-configuration Hartree–Fock methods. The energy of a single determinant wave function

$$\Phi_K = 1 / \sqrt{N!} \|\varphi_1(1)\varphi_2(2) \dots \varphi_N(N)\| \quad (14.32)$$

constructed on a set of N orthonormal spin-orbitals

$$\langle \varphi_i | \varphi_j \rangle = \delta_{ij} \quad (14.33)$$

is

$$E = \sum_i I_i + \sum_i \sum_{j < i} J_{ij} - \sum_i \sum_{j < i} K_{ij} \quad (14.34)$$

where I_i is the kinetic energy of electron i plus attraction by all nuclei A ; J_{ij} is the Coulomb repulsion between electrons i and j ; and K_{ij} is the exchange energy of electrons i and j with the same spin.

According to the variation principle, the energy minimum is obtained when

$$\frac{\partial E}{\partial \varphi_i} = 0 \quad \text{and} \quad \partial \langle \varphi_i | \varphi_j \rangle = 0 \quad (i, j = 1, 2, \dots, N).$$

In order to help differentiate the components of the energy, it is appropriate to define formal Coulomb J_j and exchange K_j operators by their action on a current function

$$J_j(\mu)\phi(\mu) = \left[\int \frac{\varphi_j(\nu)\varphi_j(\nu)d\tau_\nu}{r_{\mu\nu}} \right] \phi(\mu) \quad (14.35)$$

$$K_j(\mu)\phi(\mu) = \left[\int \frac{\varphi_j(\nu)\phi(\nu)d\tau_\nu}{r_{\mu\nu}} \right] \varphi_j(\mu). \quad (14.36)$$

After some manipulations one obtains the Hartree–Fock equation

$$\left[h + \sum_j (2J_j - K_j) \right] \varphi_i = e_i \varphi_i \quad (14.37)$$

which is a function of the φ_i orbitals themselves. This equation is therefore an iterative equation of the form

$$F\phi = e\phi \quad (14.38)$$

where F is called the Fock operator and describes the motion of the system of electrons. The usual way of numerical resolution is to project the general expression given above on a *finite* basis of atomic orbitals X_p ($p = 1, 2, \dots, m$) with $m \geq N$. After integration of the Hartree–Fock equation over the atomic basis X_p one obtains the Hall–Roothaan equation

$$\sum_p \sum_q c_{pi} c_{qi} \int X_p(r) F(r) X_q(r) dv = e_i \sum_p \sum_q c_{pi} c_{qi} \int X_p(r) X_q(r) dv. \quad (14.39)$$

The coefficients of the molecular orbitals can then be determined by solving the system of linear equations

$$\sum_q c_{qi} (F_{pq} - e S_{pq}) = 0 \quad (14.40)$$

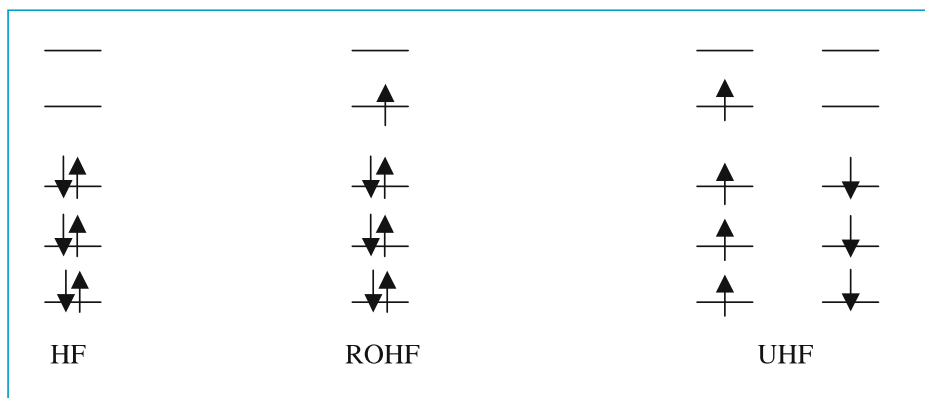


Fig. 14.4. Repartition of electrons within orbitals in single-configuration methods. In the restricted open-shell Hartree–Fock method (ROHF) all levels are occupied by two electrons except the open-shells that contain the unpaired electrons (Roothaan 1960). In the unrestricted HF methods (UHF), there is no longer a spatial constraint for electrons with different spins

where F is the Fock operator expressed in the atomic basis and

$$S_{pq} = \int X_p(r)X_q(r)dv \quad (14.41)$$

the overlap integral between functions X_p and X_q . For this system to have non-trivial solutions ($c_{pi} \neq 0$), the orbital energy e has to be the solution of the algebraic equation

$$\text{Det} |F_{pq} - eS_{pq}| = 0. \quad (14.42)$$

Through this process, one obtains m molecular orbitals φ_i . According to the Aufbau principle, only those of lower energy are occupied in the ground electronic state (Roothaan 1951; Hall 1951). The occupation scheme of the most widespread Hartree–Fock (HF) methods is presented in Fig. 14.4.

The specificity of UHF methods lies in a formalism with two different equations, one for electrons with α spin and another for electrons with β spin. The spatial relaxation thus obtained makes possible an economical access to magnetic properties while maintaining the simplicity of an independent particle model (Berthier 1954; Pople and Nesbet 1954).

b) Configuration Interaction Methods

By fixing the coefficients c_{pi} of the atomic orbitals in the molecular orbitals, one obtains the category of configuration interaction methods (CI methods)

$$\{c_{pi}\} \text{ fixed in all } \varphi_i\text{'s} \quad \text{and} \quad \frac{\partial E}{\partial d_k} = 0.$$

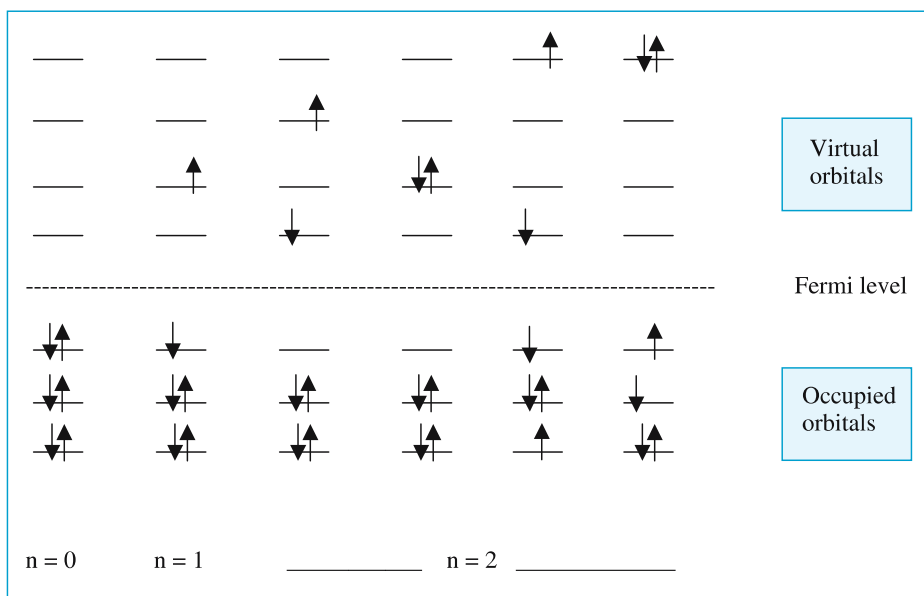


Fig. 14.5. Distribution of electrons within orbitals in CISD. The ground state determinant is defined by ($n = 0$), singly-excited determinants by ($n = 1$) and doubly-excited determinants by ($n = 2$)

For example by taking the Hartree–Fock orbitals φ_i of a preceding calculation, it is possible to generate a series of Slater determinants that can be used as a basis of configurations. The general procedure is to replace a number of spin-orbitals that are occupied in the ground state by an equal number of vacant orbitals. These orbitals, called virtual orbitals, are obtained in the SCF treatment as the orthogonal complement of the basis of the occupied functions. According to the number n of replacements, the methods are known under the following acronyms: CIS ($n = 1$), CISD ($n = 1, 2$), CISDT ($n = 1, 2, 3$) and CISDTQ ($n = 1, 2, 3, 4$). The electronic distributions used in CISD calculations (Nesbet 1955) are illustrated in Fig. 14.5.

The most widely used method for a first approximation to transition energies and the determination of excited states geometries of medium size molecules is CIS. Methods like CISD, CISDT and CISDTQ require much more computational effort; they are essentially applied for quantitative studies of energy differences (electronic transitions, conformational analysis) and calculation of observables for predictions and comparisons with experiments.

c) Multiconfigurational Self-Consistent Field Methods

By relaxing both the coefficients of the atomic orbitals in the molecular spin-orbitals and the coefficients of the configuration expansion (Hinze and Roothaan 1967),

$$\{c_{pi\text{relaxed}}\}\text{and}\{d_{K\text{relaxed}}\} \quad \frac{\partial E}{\partial c_p} = 0 \text{ and } \frac{\partial E}{\partial d_K} = 0$$

one obtains the class of multiconfigurational self-consistent field methods (MCSCF).

The quality of an MCSCF calculation depends critically on the choice of the basis of configurations, choice that has to be limited because of the complexity of the calculations. In contrast to Hartree–Fock calculations that are entirely defined by the number of electrons and the atomic basis set, MCSCF calculations can only be defined by the list of configurations Φ_K taken into account. A category of such methods, referred to as CASSCF (complete active space SCF) is widely used in today’s calculations. The variational space associated with these methods, represented by $\{m\}^N$, contains all possible electronic configurations (satisfying spatial symmetry and spin requirements) that can be generated by distributing the N active electrons within the selected m active orbitals.

As previously shown for SCF orbitals, MCSCF orbitals can also be used as a basis for the construction of new electronic configurations in view of forthcoming CI treatments. These calculations are defined by complementary distributions of the electrons within orbital spaces referred as internal space (the MCSCF space $\{m\}^N$) and external space. The external space is formed of the virtual orbitals $\{v\}$ in which electrons are promoted from the internal space. Promotion of one electron leads to first-order CI (FOCI), promotion of two electrons gives second order CI (SOCI), which is the most advanced level of configuration interaction that can be done for chemical systems with few heavy atoms (Fig. 14.6).

At this level, an important notion is that of valence space. It is the orbital space spanned by the orbitals occupied in the ground state of the atoms constituting the molecule. Transposed at the molecular level, the valence space is composed of the bonding and antibonding orbitals to which are added all electron lone pairs and, when present, all orbitals containing unpaired electrons. In other words, it is the orbital space where chemistry takes place, the space where bonds are broken and formed. Such partitioning does not concern the core electrons formed by the inner shells of the heavy atoms that are not involved in chemical reactions. It can also be seen that, if it is possible to identify the reactive part of a molecule, or the functional group that is responsible for an electronic transition, then the numerical simulation can be focused on this particular point; the remaining electrons that essentially behave as spectators can easily be handled at a lower level of theory.

The complete configuration interaction involving all possible configurations that can be generated for a given basis of orbitals (Full CI) is the only way to recover the entire correlation contribution. It can only be done for small systems with a limited number of electrons. SOCI treatments are the most advanced variational calculations that can be achieved for polyatomic systems. These calculations converge towards the FCI as the dimension of the internal space increases.

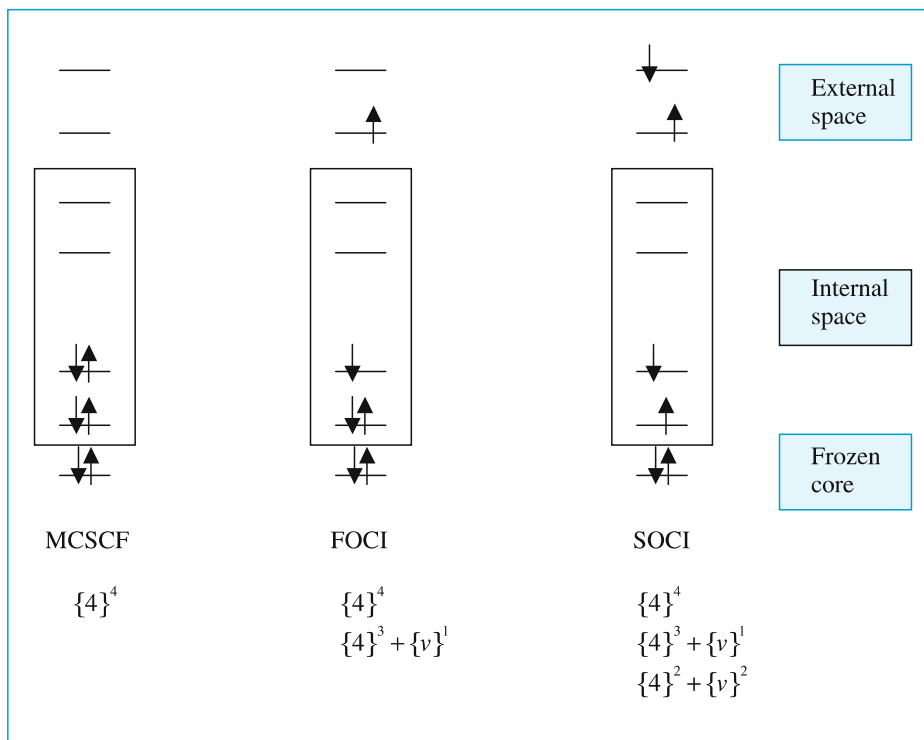


Fig. 14.6. Electronic distributions within the variational spaces of MCSCF-CI methods. All possible distributions of the electrons in the orbitals of the internal space (*box*) are taken into account

14.1.2.2 Perturbational Methods

Although very simple at the conceptual level, the variational CI is practically difficult to push to the limit due to the extremely large number of configurations that enter the calculation. Before considering the various possibilities to select a limited number of *significant configurations*, it must be realized that the overall variational process amounts to perturbing the ground state configuration by excited configuration and reciprocally. In this context, it is worth re-formulating the CI problem, qualitatively and quantitatively, in terms of the concepts of perturbation theory (Lennard-Jones 1930).

In this approach, the total Hamiltonian H is partitioned into two parts: a zeroth-order part, H^0 , whose eigenfunctions, the Φ_K and energies $E_K^{(0)}$ are known, and a perturbation V . The idea is to use the Φ_K s as a basis to expand the solutions of the exact Hamiltonian and design a procedure to systematically improve the eigenfunctions and eigenvalues of H^0 so that they become closer and closer to the eigenfunctions and eigenvalues of H . It can be done by introducing an ordering parameter λ so that

$$H = H^0 + \lambda V \quad \begin{array}{l} \lambda = 0 \text{ unperturbed system} \\ \lambda = 1 \text{ perturbed system.} \end{array}$$

The principle is to expand the exact eigenfunctions and eigenvalues in a Taylor series in λ

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots \quad (14.43)$$

$$\Psi = \Psi^{(0)} + \lambda \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \dots \quad (14.44)$$

Reporting these expressions in $H\Psi = E\Psi$ and identifying the powers in λ , one obtains

$$\begin{aligned} E_I^{(1)} &= \langle \Psi_I^{(0)} | V | \Psi_I^{(0)} \rangle = \text{first-order correction to the energy} \\ \Psi_I^{(1)} &= \sum_{K \neq I} \frac{V_{KI}}{E_I^{(0)} - E_K^{(0)}} = \text{first-order correction to the wave function} \\ E_I^{(2)} &= \sum_{K \neq I} \frac{|V_{KI}|^2}{E_I^{(0)} - E_K^{(0)}} = \text{second-order correction to the energy.} \end{aligned}$$

Successive corrections to the energy are much more easily obtained than corrections to the wave function since the knowledge of the n th order function makes it possible to evaluate the energy up to $2n + 1$ order.

It was Möller and Plesset (1934) who proposed to apply this perturbation theory to the problem of electronic correlation using as a perturbation the correlation potential defined as the difference between the exact electronic Hamiltonian and the Hartree–Fock operator

$$V = H - H_{\text{HF}}. \quad (14.45)$$

The basis functions used in this formulation are the eigenfunctions Φ_K of H_{HF} that are built from the Hartree–Fock orbitals φ_i . The energies (the diagonal terms) are summation of one-electron energies and it follows that energy differences are differences of one-electron orbital energies (the e_i from the Hartree–Fock calculation).

The energetic terms necessary to the evaluation of the perturbation corrections are the diagonal $\langle \Phi_I | V | \Phi_I \rangle$ and non diagonal elements $\langle \Phi_I | V | \Phi_J \rangle = \langle \Phi_I | H | \Phi_J \rangle$ between the different configurations of the basis of determinants Φ_K .

According to the order $[n]$ of the perturbation, the configurations that are active in the process are characterized by an increasing number of electronic excitations $i = 1, 2, 3, 4, 5, \dots$. However, these configurations can interact only if they differ by less than two spin-orbitals. In all the other cases, the interaction is equal to zero. The structure of the CI matrix is illustrated in Table 14.2.

Table 14.2. Order of perturbation [n] when the electronic configurations enter the calculation

Mono excited	Di excited	Tri excited	Quadri excited	Penta excited
0	[2]	0	0	0
[5]	[4]	[5]	0	0
	[3]	[4]	[4]	0
		[5]	[5]	[6]
			[5]	[6]

In the Möller–Plesset formalism (1934)

$$\Psi_0^{(0)} = \Phi_0 \quad (14.46)$$

and

$$E^{(0)} + E^{(1)} = E_{\text{SCF}} . \quad (14.47)$$

In other words, the total SCF energy is equal to the energy corrected to first-order. The first-order correction to the wave function is

$$\Psi_0^{(1)} = - \sum_{K \neq 0} \frac{\langle \Phi_K | H | \Phi_0 \rangle}{E_K^{(0)} - E_0^{(0)}} \Phi_K . \quad (14.48)$$

This expression makes use of configurations $\Phi_K = \Phi_{ij}^{ab}$, where two electrons are excited from occupied orbitals φ_i and φ_j into virtual orbitals φ_a and φ_b , i.e.,

$$\Psi_0^{(1)} = \frac{1}{4} \sum_{i,j}^{\text{occ}} \sum_{a,b}^{\text{virt}} \frac{\langle \varphi_a \varphi_i | V | \varphi_b \varphi_j \rangle - \langle \varphi_a \varphi_j | V | \varphi_b \varphi_i \rangle}{e_a + e_b - e_i - e_j} \Phi_{ij}^{ab} . \quad (14.49)$$

Hence, the second-order correction to the energy is

$$E_0^{(2)} = \frac{1}{4} \sum_{i,j}^{\text{occ}} \sum_{a,b}^{\text{virt}} \frac{|\langle \varphi_a \varphi_i | V | \varphi_b \varphi_j \rangle - \langle \varphi_a \varphi_j | V | \varphi_b \varphi_i \rangle|^2}{e_a + e_b - e_i - e_j} . \quad (14.50)$$

After the second-order correction, the Möller–Plesset total energy is equal to $E_0^{\text{MP2}} = E_{\text{SCF}} + E_0^{(2)}$. Higher orders of perturbation treatments are known under acronyms MP3 and MP4.

14.1.2.3 Variation-Perturbation Methods

The idea at the origin of this class of methods is to take the advantages of both variational and perturbational approaches in a unique formalism. Two different types of methods are obtained according to which of the variational or perturbational treatment is performed first.

Very compact treatments are obtained if the configurations Φ_K that provide the most important contributions to the energy are selected according to their second-order contribution to form a configuration expansion whose coefficients d_K are optimized in the following variational calculation. The overall method is a multistep process:

- Step 0: Calculation of the Hartree–Fock wave function Φ_0 .
- Step 1: Second order perturbation treatment followed by selection of configurations Φ_K whose contribution to the energy is greater than a predetermined threshold.
- Step 3: Variational calculation of the wave function in the basis of the Φ_K s. The major inconvenience of this technique is that the selected basis of configurations may fluctuate with geometry along a reaction path. This well-known drawback may be rectified by an iterative procedure in which the configuration list of Φ_K is updated in successive variation-perturbation cycling as follows. In this way, configurations whose contribution to the energy is smaller and smaller are progressively introduced into the calculation.
- Step 4: Second-order perturbation treatment of the multiconfigurational wave function obtained in step 3 and selection of configurations whose contribution to the energy is greater than a threshold an order of magnitude smaller than in step 1.
- Step 5: Variational calculation of the wave function in the extended basis of the Φ_K s.

The process is iterated over steps 4 and 5 until convergence is reached. It is the principle at the origin of the CIPSI method (Huron et al. 1973).

In the other type of methods, it is the MCSCF wave function that is improved by using second-order perturbation theory. When the variational treatment is of CASSCF type, the method is known as CAS-PT2 or MC/P for any general definition of the MCSCF active space.

By extension, one may consider that methods like coupled cluster (Bartlett 1981) in which the variational principle is applied per type of energy contributions up to doubly-excited states (e.g., CCSD) and followed by a perturbation treatment to account for the effect of higher excitations (e.g., CCSD(T)) belong to the same family.

14.1.3 DFT Methods

Density functional theory (DFT) is a totally different approach compared to the methods discussed previously that are based on the determination of the wave function. DFT relies on the fact that it is not necessary to know the wave function $\Psi(r_1, r_2 \dots r_N)$ of the N electrons of the system to obtain the energy and properties of the electronic ground state; only the knowledge of the electronic density $\rho(r)$ is mandatory. Contrary to wave function methods where an exact Hamiltonian operator is used and approximations made to the wave function, DFT methods use an approximate Hamiltonian but determine the electronic density exactly.

In the preceding section, wave function methods have been presented in the order of increasing complexity, i.e., considering progressive tuning of the approximations to the wave function; in the following section DFT methods will also be presented in the order of increasing complexity, i.e., considering progressive tuning of the Hamiltonian operator.

14.1.3.1 Hohenberg and Kohn Theorems

In quantum mechanics, the wave function of an electronic system in its ground state and the associated electronic density are uniquely determined by the number N of electrons and the external potential to which the system is submitted (the electron-nucleus attraction). The reverse proposition was proved in 1964. It is the first Hohenberg and Kohn theorem (1964): the external potential is determined, to a trivial additive constant, by the electronic density.

In other words, there is a biunivocal correspondence between the external $V(r)$ and the energy of the electronic ground state. However, the fact that the external potential $V(r)$ is unique for a given density $\rho(r)$ does not imply that such an external potential exists for any electronic density. If this potential does exist, then the associated density is said to be “ V -representable”. In the same way, an electronic density that can be obtained from some antisymmetric wave function (for example a Slater determinant or a combination of determinants), is said to be “ N -representable”.

It can be seen that the knowledge of the exact electronic density is sufficient to determine the number N of electrons together with the charges and positions of the nuclei. In other words, the exact Hamiltonian is known and consequently any other property of the system can be evaluated.

The second Hohenberg and Kohn theorem states the variational principle for DFT methods: for any trial electronic density such that $\rho(r) \geq 0$ and $\int \rho(r) dr = N$, then $E_v[\rho] \geq E_0$.

In this expression, E_0 is the exact energy (not known) and index v means that the energy is V -representable.

It must be stressed here that, if different electronic densities lead to different energies for the ground state, the mathematical relation connecting the energy to the density is unknown. It is the purpose of actual developments in density

functional theory to bridge this gap by building the appropriate mathematical tools.

a) *Functions and Functionals*

A *function* $F[x]$ is a rule which, when applied to a set of variables (coordinates for example) gives a number. The wave function and the electronic density are such functions. A *functional* $F[f]$ is a rule which, when applied to a function that itself depends on a set of variables, gives a number. The energy, $E_{\text{Ae}}[\rho]$, which depends on the density ρ (or wave function) is a functional, according to this definition.

The *extremum* of a functional is given by the relation

$$\frac{\partial F[f]}{\partial f} = 0 \quad (14.51)$$

obviously the analogue of a function extremum.

b) *Expression of the Total Energy*

The total electronic energy for a system formed of A nuclei with charges Z_A and N electrons can be written as

$$E_{\text{Ae}}[\rho] = T[\rho] + V_{\text{Ae}}[\rho] + E_{\text{ee}}[\rho] \quad (14.52)$$

where

$$\begin{aligned} T[\rho] &= \text{kinetic energy of the electrons} \\ V_{\text{Ae}}[\rho] &= \text{electron-nucleus attraction energy} \\ E_{\text{ee}}[\rho] &= \text{electron-electron interaction energy.} \end{aligned}$$

As before, the repulsion between nuclei, V_{AA} , is a constant within the Born-Oppenheimer approximation; it is a posteriori added to the electronic energy to obtain the total energy of the system.

The attraction between the electrons and the nuclei is given by

$$E_{\text{Ae}}[\rho] = \sum_A \int \frac{Z_{\text{SA}}\rho(\mathbf{r})}{|R_A - \mathbf{r}|} d\mathbf{r} \quad (14.53)$$

As in wave function based methods, the problem is the description of the electron-electron interaction. In DFT it is expressed by

$$E_{\text{ee}}[\rho] = J[\rho] + (E_{\text{ee}}[\rho] - J[\rho]) \quad (14.54)$$

where $J[\rho]$ is the Coulomb repulsion given by

$$J[\rho] = \frac{1}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' . \quad (14.55)$$

In these equations, $T[\rho]$ and $(E_{ee}[\rho] - J[\rho])$ describe the kinetic energy of the electrons and the part of the interaction between electrons that is not a pure Coulomb term. These quantities have a quantum mechanical origin and there is no classical expression to be used as an alternative. The main objective in DFT is the development of more accurate descriptions of these two terms in order to be able to determine the minimum of the energy functional and obtain the associated electronic density together with the total energy of the system.

14.1.3.2 Thomas–Fermi–Dirac Model

The first description of an atom that does not rely on a wave function is due to Thomas and Fermi. The basic idea is that the electronic cloud can be taken as a uniform gas of independent electrons, namely a gas of non-interacting electrons. In this context, the kinetic energy of the electrons is a function of the electron density

$$T_{TF}[\rho] = C_x \int \rho^{5/3}(r) dr \quad (14.56)$$

where

$$C_x = \left(\frac{3}{4}\right) \left(\frac{3}{\pi}\right)^{1/3} = 2.871 \text{ a.u.}^5$$

In this model, the energy $T_{TF}[\rho]$ is calculated supposing that each infinitesimal element of volume contains an electronic density that is locally uniform. Extending the approximation to functionals other than the kinetic energy leads to what is referred to as the local density approximation or LDA.

To improve this model Dirac introduced a supplementary term, called the exchange term, which is given by

$$K_D[\rho] = -C_x \int \rho^{4/3}(r) dr \quad (14.57)$$

with

$$C_x = \left(\frac{3}{4}\right) \left(\frac{3}{\pi}\right)^{1/3}. \quad (14.58)$$

This type of method, even the Thomas–Fermi–Dirac model presented above, are not adapted to molecular calculations since the approximation of the electron cloud by a uniform gas of electrons cannot really account for the localized chemical bonds. All the improvements proposed for these models are far from providing acceptable results when compared to the wave function methods.

⁵ In chemistry, a.u. means atomic unit. For energy, 1 a.u. = 27.21 eV (twice the energy of the hydrogen atom); for bond lengths, 1 a.u. = 0.529177 Å (Bohr radius).

14.1.3.3 Kohn–Sham Equations

It is by analogy with wave function methods that Kohn and Sham proposed to introduce the notion of orbital in the DFT formalism (Kohn and Sham 1965). In this new framework, the kinetic energy explicitly depends of the orbitals

$$T[\rho] = \langle \Psi | \sum_{i=1}^N -\frac{1}{2} \nabla_i^2 | \Psi \rangle \quad (14.59)$$

where Ψ is the ground state wave function. The next step, which the majority of the actual Kohn–Sham developments rely on, is a direct consequence of the Hartree–Fock independent particles model when adapted to the DFT formalism.

Let us suppose that a single determinant wave function Ψ_s is able to reproduce the electronic density of the exact wave function. Let us suppose in addition that the electrons in the φ_i orbitals do not interact. The kinetic energy is given by

$$T_s[\rho] = \langle \varphi_i | \sum_{i=1}^N -\frac{1}{2} \nabla_i^2 | \varphi_i \rangle . \quad (14.60)$$

The total energy functional can then be rewritten in the form of an exact expression referred to as the Kohn–Sham functional

$$E[\rho] = T_s[\rho] + J[\rho] + E_{Ae}[\rho] + E_{xc}[\rho] \quad (14.61)$$

in which $E_{xc}[\rho]$ is called the exchange–correlation functional

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (E_{ee}[\rho] - J[\rho]) . \quad (14.62)$$

This particular functional has the property of containing all the terms of density functional theory whose analytical expressions are unknown. The exchange–correlation functional contains a part of the kinetic energy, $T[\rho] - T_s[\rho]$, that accounts for the difference between the kinetic energy of the supposedly independent electrons and that of the interacting electrons. It is worth mentioning that such a contribution is not part of the correlation energy calculated in wave function based methods.

The components that provide the major part of the total energy functional, i.e., $T_s[\rho]$, $J[\rho]$ and $E_{Ae}[\rho]$ can be evaluated exactly using the electronic density, which is easily obtained in the orbital model

$$\rho = \sum_{i=1}^N |\varphi_i|^2 . \quad (14.63)$$

The specifics of the various DFT methods follow from the expressions given to the $E_{xc}[\rho]$ functional. They may be considered as levels of theory by analogy with wave function methods.

Introducing an orbital representation changes the theoretical scheme. Instead of minimizing the energy with respect to variations of the electronic density, the problem is transformed into a minimization with respect to variations of the orbitals. The same approach that allowed us to derive the Hartree–Fock equations in wave function methods now leads to the similar Kohn–Sham equations in DFT

$$h_{\text{KS}}\varphi_i = \varepsilon_i\varphi_i \quad (14.64)$$

with

$$h_{\text{KS}} = -\frac{1}{2}\nabla_i^2 + V_{\text{eff}} \quad (14.65)$$

where V_{eff} is referred to as the Kohn–Sham effective potential. In this equation h_{KS} is a one-electron operator whose eigenfunctions φ_i are the Kohn–Sham orbitals. As in Hartree–Fock methods, the highest eigenvalue has the meaning of an ionization potential. No particular significance can be attached to the other eigenvalues.

The Kohn–Sham effective potential

$$V_{\text{eff}} = v(r) + \int \frac{\rho(r')}{|r-r'|} dr' + v_{xc}\{\rho(r)\} \quad (14.66)$$

is composed of

$$\begin{aligned} v(r) &= \text{external potential (nuclear attraction)} \\ \int \frac{\rho(r')}{|r-r'|} dr' &= \text{Coulomb repulsion potential} \\ v_{xc}\{\rho(r)\} &= \text{exchange-correlation potential.} \end{aligned}$$

The first step of any calculation is to take a reasonable approximation of $E_{xc}[\rho]$, then to determine the $v_{xc}\{\rho(r)\}$ potential by analytical differentiation $\partial E_{xc}/\partial\rho$.

The process for solving the Kohn–Sham equations is similar to that used for the Hartree–Fock equations:

- Choice of a trial set of orbitals
- Calculation of the trial density
- Construction of the effective Kohn–Sham potential
- Diagonalization of the $\langle\varphi_i|h_{\text{KS}}|\varphi_j\rangle$ matrix

The orbitals obtained in this process are taken as trial orbitals for the next iteration and the procedure is repeated until convergence. The value of the total energy

$$E = \sum_{i=1}^N \varepsilon_i - J[\rho] + E_{xc}[\rho] - \int v_{xc}\{\rho(r)\}\rho(r)dr \quad (14.67)$$

can then be obtained together with the molecular properties.

As in wave function methods, a linear combination of atomic orbitals is used to represent the molecular orbitals (LCAO approximation). They are Gaussian type functions or exponential functions leading to equations formally equivalent to Roothaan equations. In the same way, the extension of the atomic basis set allows numerical convergence of the results towards the exact solution.

14.1.3.4 Local Density Approximation (LDA and LSDA)

Within this approximation, the unknown exchange-correlation functional is approximated by the exact exchange-correlation functional of a uniform electron gas

$$E_{xc}^{\text{LDA}}[\rho] = \int_V \rho(r) \varepsilon_{xc}(\rho) dV \quad (14.68)$$

with

$$\varepsilon_{xc}(\rho) = \varepsilon_x(\rho) + \varepsilon_c(\rho) \quad (14.69)$$

where $\varepsilon_x(\rho)$ is the exact exchange energy given by Dirac and $\varepsilon_c(\rho)$ the correlation energy obtained by adjustment on a uniform electron. The expression proposed by Vosko, Wilk and Nusair (1980) is the most commonly used (VWN).

Total energies and interaction energies provided by the LDA approach are generally overestimated. Bonding energies are too strong and interatomic distances are too short. In the case of open-shell systems, the results are far from satisfactory. A net improvement consists in considering the densities of α and β electrons separately as in unrestricted Hartree–Fock methods. This spatial relaxation of the spin densities, referred to as the local spin density approximation improves the results significantly, but is too uncertain from one system to another.

14.1.3.5 Generalized Gradient Approximation (GGA)

In the methods that rely on the LDA approximation, it is postulated that the electronic density is constant within every infinitesimal volume element dV . This hypothesis is no longer enforced in the methods based on the generalized gradient approximation. In these GGA methods, it is postulated instead that both the exchange and the correlation not only depend on the density but also on the gradient of this density. This way of proceeding is often considered as a non-local approach.

Apart from the dependence on the density derivatives, the exchange-correlation functionals are quite similar in the LDA and GGA methods. The various expressions that have been proposed can be classified into two categories. The first one is that of the corrections or new formulations of the exchange part of the functional, as for example the PW86 (Perdew and Wang 1986) and B88 (Becke

1988); in the second one are implementations of new correlation functionals, e.g., PW91 (Perdew and Wang 1991) or LYP (Lee, Yang and Parr 1988) in which some parameters have been adjusted to reproduce experimental data of the helium atom.

In its present form, the GGA approximation provides theoretical values that compare favorably to the first levels of ab initio correlated methods. Possibilities to improve the actual status are being investigated following two orientations:

1. The first consists in the development of empirical functionals; the idea is to adjust the functionals to the properties to be calculated. It looks like a return towards semi-empirical problematics with the advantage of being certain to get better numerical values for a specific observable in a given context and the inconvenience of losing the general aspect of the concept of functional.
2. The second is a non-empirical approach. In this case, the parameters of the functionals are not adjusted to some experimental data, but constrained to fulfill specific mathematical requirements.

A more recent direction of research is to find ways to include the dependence on the second derivative of the density in addition to the first derivative already considered. It is actually referred to as the meta-GGA approach.

14.1.3.6 Adiabatic Connection and Hybrid Methods

In this type of method, the electronic repulsion is expressed as a function of a tuning parameter λ such that, for $\lambda = 1$ the interaction is *on*, and for $\lambda = 0$ the interaction is *off*.

The exchange-correlation functional can then be written as a function of the λ parameter (continuous)

$$E_{xc} = \int_0^1 (\langle \Psi_\lambda | E_{ee} | \Psi_\lambda \rangle - J[\rho]) d\lambda \quad (14.70)$$

where Ψ_λ is the exact wave function for a given λ . In the half and half approach proposed by Becke, the integral term is replaced by the average of the two terms corresponding to $\lambda = 0$ and $\lambda = 1$.

- For $\lambda = 0$ one has the equivalent of the Hartree–Fock approximation with independent electrons. In this case, E_{xc} does not contain any correlation energy, but the exchange energy is taken into account. The wave function, indeed, being represented by a single determinant built on the Kohn–Sham orbitals, the exchange energy can be exactly evaluated as in Hartree–Fock theory.
- For $\lambda = 1$, the exchange-correlation functional E_{xc} can be taken as the GGA functional.

The qualification of hybrid methods, given to this type of developments means that the two approaches of wave function and DFT are mixed for the calculation of E_{xc} .

Another well-known and fashionable hybrid functional is B3LYP. Proposed by Becke (1993), it is actually employed for all sorts of applications in chemistry, from energy determinations to magnetic observables. Three different exchange functionals are combined, E_x^{LDA} , $E_x^{\text{exact(HF)}}$ and E_x^{B88} together with two correlation functionals E_c^{LYP} and E_c^{VWN} . The complete form of the exchange-correlation functional is now expressed by

$$E_{xc}^{\text{B3LYP}} = (1 - a)E_x^{\text{LDA}} + aE_x^{\text{exact(HF)}} + b(E_x^{\text{B88}} - E_x^{\text{LDA}}) \quad (14.71)$$

$$+ E_c^{\text{VWN}} + c(E_c^{\text{LYP}} - E_c^{\text{VWN}}). \quad (14.72)$$

The values of the a , b and c parameters are 0.2, 0.7 and 0.8, respectively. They have all been adjusted, empirically, so as to reproduce best the experimental data on geometries and energetics of series of molecules containing first and second row atoms.

14.1.4 Wave Function Versus DFT Methods

The choice of the method to be employed results from a compromise critically dependent on the two dimensions of the problem, the chemical dimension and the computational dimension.

The first dimension is given by the number of electrons in the system. For the calculation of chemical properties or reaction profiles, this dimension can be reduced, either by freezing the electrons that behave as spectators or by using effective potentials for core electrons.

The second dimension depends on the number of atomic functions used for the determination of the molecular orbitals and on the number of configurations (given by the number of excitations) that have to be generated to describe the wave function accurately. One has to keep in mind that the dimension of the atomic basis set and the number of configurations that can be generated are not independent parameters in *ab initio* post Hartree–Fock methods. For a result to be reliable (preferably with an estimation of the error bar) it is necessary to reach *convergence* of the numerical values with the increase in the size of the atomic basis *and* with the increase in the size of the configuration expansion. In a general way the calculation must be well-balanced between the size of the basis set and the level of excitation in the configuration interaction (Fig. 14.7).

One of the great advantages of DFT compared to wave function methods is that the electron density depends on the same three variables (coordinates X , Y , Z) whatever the dimension of the system. However, there is no sequence of calculations that can be considered equivalent to the convergence of *ab initio* methods with the size of the configuration expansion. The electronic correlation

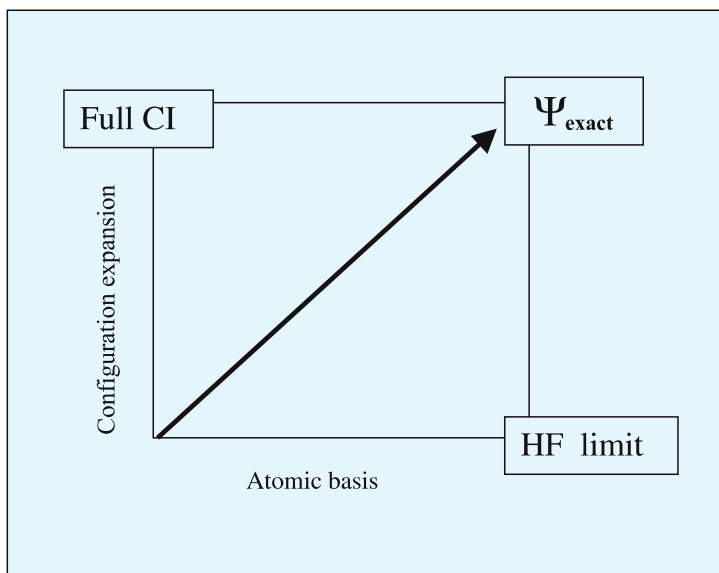


Fig. 14.7. Looking for the exact wave function in ab initio methods

being taken into account via the exchange-correlation functional, the reasonable solution is to look for the functional best adapted to the problem. Most often, one can rely on a more or less empirical knowledge drawn from an increasing body of data on DFT calculations found in the literature; otherwise, it must be done by trial and error.

Computational constraints become rapidly a limiting factor. The computational cost increases as N^3 for Hartree–Fock or DFT treatments; it increases as N^5 for CISD and can reach as high as N^6 or N^7 for more sophisticated coupled cluster calculations. In current variational calculations, it is quite common to determine several of the lowest eigenvalues and corresponding wave functions of interaction configuration matrices of dimension about $10^6 \times 10^6$ or even larger.

14.2 Applications of Quantum Chemistry

Practically, it is the nature of the problem to be solved that decides the type of simulation to be made. Two possibilities are to be considered:

1. The main goal of the study is to reproduce an experimental result or observation in order to get an explanation or rationalization of the facts: it is the domain of *postdictions*. A semi-quantitative agreement is generally sufficient.
2. The main goal of the study is to provide an alternative to an experiment: it is the domain of *predictions*. No experimental data being available, it is

necessary to design a series of calculations converging towards the exact wave function.

In fact, the precision that can be obtained greatly depends on the observable considered. One has to distinguish between energetic properties associated with potential energy surfaces and spectroscopic properties implying different energy levels. The main difficulty in the later case is to give an equivalent treatment to all states involved in the process. In other words one has to arrange that the computational uncertainty is the same for all levels, so that the differences in energies are correct. How numerical simulation can respond to the demands of exobiology is illustrated in Table 14.3.

The calculation of spectral signatures and reaction processes involving small systems will be done at the highest possible level. For example, the prediction of the rotational spectra in order to discriminate between all the possible isomers of general formula [HPCO], or the determination of the reaction paths of the system [N + CH₃] need what is called state of the art calculations. Large-scale configuration interaction methods with extended basis sets will be employed.

Precision, here, is the main concern, because quantum chemical calculations serve to provide the starting data for a posteriori collision treatments. These treatments are very sensitive to small variations in the shapes of the potential surfaces. High precision is critical for the determination of the values characteristic of saddle points and second derivatives of the energy with respect to the displacements of the nuclei; small variations may have drastic consequences on reaction rates and branching ratios that are the quantities needed to understand interstellar chemistry.

Table 14.3. General possibilities and actual limitations of the calculations of observables

Precision requested (by the astrophysical community)		Response (of theoretical chemistry)
Rotational constants	<1%	Possible
Dipole moment	5%	Possible
Vibration frequencies	<1%	Almost
Absolute intensities	10%	Not yet
Electronic transitions	<0.2eV	Still work to be done
Proton affinity	1 kcal/mol	Possible
Isomerization energy	1 kcal/mol	Possible
Dissociation energy	1 kcal/mol	Almost
Physisorption energy	0.1 kcal/mol	Possible
Ionization energy	5 kcal/mol	Not yet

For medium-size compounds, say up to ~ 50 atoms, as polycyclic aromatic hydrocarbons (PAH), the methods used will be essentially MP2 or DFT (DFT calculations will be preferred as the size of the target system increases). It is the most economical way to take into account correlation effects that are important in the proper description of infrared spectra, for example. In general, the raw values are calibrated on well-known experimental data. Various extrapolation techniques can also be employed for series of molecules as shown below in examples of linear chains.

For very large systems, on the order of hundreds of atoms, numerical simulations are based on semi-empirical methods and well-parameterized effective potentials that are used in the frame of molecular mechanics.

In spite of the different fashions and schools, it should be stressed that there is no universal miraculous method. Every different problem should be handled with the appropriate method (when existing), if not, the proper method should be invented (Fig. 14.8).

In this context, calculation strategies must be tested (or calibrated) on reliable and carefully selected experimental data and then applied to unknown species. It can be seen that postdictions cannot be separated from predictions

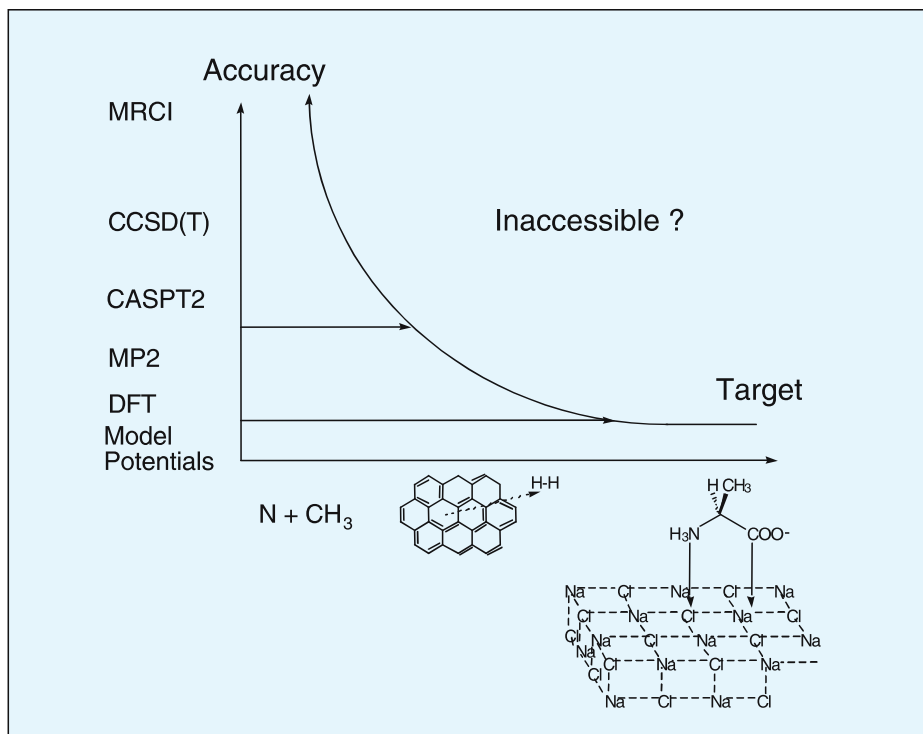


Fig. 14.8. One problem, one method

in the sense that the former have to be done and understood for the latter to be achieved in the best possible ways.

14.2.1 Radio Millimeter Observations

Identification of a molecule in space relies on a perfect match between the observed lines and the spectrum of the same species in the laboratory. It is something difficult to realize, especially for unstable species. Generally indeed, these species are formed together with a number of other molecules that may be isomers or even compounds with completely different chemical formulas. The role of numerical simulation is to propose theoretical spectra that are precise enough to make possible an interpretation of the laboratory spectra and extract the spectrum of the target molecule from the mixture of experimental values for comparison with the spatial observations.

14.2.1.1 Calibration of Molecular Calculations

Prediction of rotational constants with enough accuracy to make it possible to identify an unknown molecule with certainty is at the very limits of present computational chemistry. The major difficulty comes from the fact that the calculations are made within the rigid rotor approximation, thus leading to constants A_e , B_e and C_e at the equilibrium geometry, whereas the quantities needed to reproduce the radio signatures (Townes and Schawlow 1975) are constants A_0 , B_0 and C_0 (taking into account the motion of the nuclei). This problem, however, is not without solutions.

It has been shown that the raw values given by quantum chemistry calculations, whose accuracy is on the order of $\sim 1\%$ can be corrected so as to obtain the 0.3% precision that is generally considered as the upper limit required for a possible identification. All the calibration procedures rely on some combination of the calculated values with experimental data on well-known related compounds. Different approaches are used according to the type of target molecule.

A first scheme of calibration is applicable to the rotational constants themselves when the laboratory spectra of isoelectronic systems of comparable masses have already been determined. The conjecture is that the ratio $B_{\text{iso}}^{\text{Theory}}/B_{\text{iso}}^{\text{Exp}}$, where B_{iso} may be any of the constants A , B or C of the reference isoelectronic molecule, does not depend on the molecule at a given level of theory. Such an approach has given remarkable results for the isoelectronic series HOCN, HN₂, HNC and HOCO⁺ for which it led to identification of the positive ion (De Frees et al. 1982).

Another strategy of calibration consists in defining bond length and bond angle increments by confronting calculated geometrical parameters to those deduced from experiments on small systems of similar structure. It is the technique used for HOCO⁺ and the series of sulfur analogues, HOCS⁺, HSCO⁺ and HSCS⁺ (Taylor and Scarlett 1985).

A different approach is that of extrapolation methods. It is particularly well-adapted to the prediction of rotational constants in series of homologous molecules whose leading terms are known. The basic idea is to determine an analytical relation that links the calculated constants to the experimental values. The analytical form generally employed presents a quadratic dependence (De Fries and McLean 1989)

$$B_0^{\text{Observed}} = a + bB_e^{\text{Theory}} + c(B_e^{\text{Theory}})^2 . \quad (14.73)$$

This expression, evaluated at a given level of theory, is then used for the calculation of the rotational constants of the next members of the series. It is the strategy that has been employed for the cyanopolyynes series: $\text{H} - (\text{C} \equiv \text{C})_n - \text{CN}$. A closely related scheme has been used for predicting the rotational constants of the long chain polyynyl, $\text{C}_n - \text{H}$ and cyanopolyynes $\text{C}_n - \text{CN}$ radicals. In these series, it has been found that the ratio $B_0^{\text{Observed}}/B_e^{\text{Theory}}$ converges toward a constant with increasing n (Pauzat et al. 1991). This result was exploited to predict the rotational constants of $\text{C}_n - \text{H}$ radicals ($n = 7-10$). These theoretical predictions supported the identification of C_7H , C_8H and C_9H in the laboratory (Travers et al. 1996; McCarthy et al. 1996a; McCarthy et al. 1996b) and the detection in space of C_7H (Cernicharo and Guélin 1996) and C_8H (Guélin et al. 1997). It was a confirmation that the limit of 0.3% for the error bar could be reached in the simulations (Table 14.4).

The examples presented in Table 14.4 show that, with proper combination of experimental values and theoretical results, it is not only possible to rationalize observations by identifying the “good candidates” but also to predict spectral signatures well before the experiments or the observations are performed. In the

Table 14.4. Rotational constants, predicted and observed, for linear radical chains

Molecule	Observation		Calculation	Prediction	$\Delta B_e\%$
	B_0 (GHz)	Year	HF B_e (GHz)	1991 [•] B_e (GHz)	
C_2H	43.675	1974	43.322		
C_3H	11.186	1985	11.067		
C_4H	4.759	1978	4.7350		
C_5H	2.395	1986	2.3732		
$\text{C}_6\text{H}^{*\#}$	1.3862	1986	1.3757		
$\text{C}_7\text{H}^{*\#}$	0.8744869	1997	0.86734	0.873 ± 0.002	0.2%
$\text{C}_8\text{H}^{*\#}$	0.5866707	1996	0.58223	0.586 ± 0.001	0.1%
$\text{C}_9\text{H}^{*\#}$	0.4132579	1997	0.40962	0.412 ± 0.001	0.3%

* Cernicharo, Guélin: observations; # Thaddeus: laboratory experiments; •Pauzat: theoretical calculations (see references in the text)

1980s, the calculations were done at the Hartree–Fock level with basis sets of average quality; it was the best that could be done at that time if all terms in the series were to be treated on the same footing. Since then, methods (and computational facilities) have greatly improved and the same studies would not be done today at the same level of theory.

14.2.1.2 Choice of a Level of Theory

The search for new molecules is sometimes confronted with a major problem, i.e., the absence of a permanent dipole of the target molecule (the intensity of the radio spectrum scales with the square of the dipole moment). Therefore, molecules like N_2 or CO_2 are not *visible* in radio. These same molecules, however, become perfectly detectable under their protonated forms. Incidentally, it should be stressed that all the positive ions identified in the ISM are protonated ions (HCO^+ , $HOCO^+$, HCS^+ , HNN^+ , $HCNH^+$ and $HCCCNH^+$).

The series of molecules with general formula OC_nO is a typical example of unobservable compounds. The first term, $HOCO^+$, formed by protonation of CO_2 is an interesting example to show the progress made by numerical simulation with the emergence of DFT methods. It can be seen in Table 14.5 that ab initio calculations are approaching the experimental values better and better as the levels of theory improve (HF, perturbative treatment MP4, or coupled cluster CCSDT).

The structures calculated using the DFT/B3LYP method provide rotational constants with an error bar on the order of $\sim 0.5\%$ in the average. The difference between theory (Cheikh and Pauzat 1999) and experiment (Bogey et al. 1984; 1986) comes from the rigid rotor approximation used in the calculations. Including vibrational effects on rotational constants is nowadays possible for small

Table 14.5. Rotational constants, observed and calculated, for $HOCO^+$ (GHz). It can be seen that the convergence illustrated in Fig. 14.6 is not yet achieved. The DFT approach is much more reliable in this case and the B3LYP formalism is well-adapted to the determination of the geometries

Models	B	C
Ab initio		
HF/6-311G**	11.0700	10.9228
MP4/6-311G**	10.5515	10.4029
CCSDT/6-311G**	10.6797	10.5269
DFT		
B3LYP/6-311G**	10.8140	10.6693
B3LYP/aug-cc-pVTZ	10.8041	10.6590
Experiment	10.7737221	10.6094435

systems. It is still impossible for molecules beyond five to six atoms and the small improvements expected (if any) are not generally worth the computational effort.

Prior to the investigation of the HOC_nO^+ series, it is better to check that the method that gives good results for the first term of the series ($n = 1$) still performs without loss of accuracy for a series of closely related molecules when the length of the chain increases. As a test case we consider the C_nO series whose nine first terms have been identified, either as singlet states (Brown et al. 1985; Ogata et al. 1995) or triplet states (Ohshima et al. 1995).

Comparison of atomic basis sets (Table 14.6) shows that the aug-cc-pVTZ basis, especially designed for ab initio correlated calculations, does not provide any improvement in DFT. This is due to the fact that the accuracy of the calculations is already high enough to be in the domain where vibrational and centrifugal corrections (neglected in these calculations) compensate the incompleteness of the treatment of correlation effects. At this level of precision, any further extrapolation or scaling procedure will, most of the time, result in a de-

Table 14.6. Rotational constants, observed and calculated, for C_nO (MHz)

Molecule*		Experiment	Theory			
			B3LYP/ 6-311G**	$\Delta B_e\%$	B3LYP/ aug-cc-pVTZ	ΔB_e
CO	S	57 635.9660(17)	58 037.22	0.7	58 154.55	0.9
C_2O	T	11 545.5970(7)	11 578.94	0.3	11 624.49	0.7
C_3O	S	4810.8864(2)	4801.812	0.2	4816.985	0.1
C_4O	T	2351.2625(2)	2349.320	0.08	2357.745	0.3
C_5O	S	1366.84709(6)	1364.760	0.15	1369.780	0.2
C_6O	T	849.75709(6)	849.2431	0.06	852.300	0.3
C_7O	S	572.94105(5)	572.6526	0.05		
C_8O	T	400.64183(8)	400.6327	0.00		
C_9O	S	293.73611(4)	293.7576	0.01		

* S: singlet (all electrons paired); T: triplet (two unpaired electrons with same spin)

Table 14.7. Spectroscopic constants (GHz) and dipole moments (Debye) of $\text{OC-CH-C}_{(n-2)}\text{O}^+$ (B3LYP/6-311G* method)

Molecule	μ	A	B	C
OC-CH-CO^+	1.86	31.458 ± 0.06	2.6963 ± 0.005	2.4835 ± 0.005
$\text{OC-CH-C}_2\text{O}^+$	3.08	35.762 ± 0.07	1.4554 ± 0.003	1.3985 ± 0.003
$\text{OC-CH-C}_3\text{O}^+$	1.73	17.732 ± 0.02	0.9314 ± 0.0009	0.8849 ± 0.0009
$\text{OC-CH-C}_4\text{O}^+$	4.00	15.121 ± 0.01	0.6197 ± 0.0006	0.5953 ± 0.0006
$\text{OC-CH-C}_5\text{O}^+$	1.70	13.691 ± 0.01	0.4215 ± 0.0004	0.4089 ± 0.0004
$\text{OC-CH-C}_6\text{O}^+$	4.55	11.343 ± 0.01	0.3095 ± 0.0003	0.3013 ± 0.0003

terioration of the results. The preliminary approach presented above leads us to adopt the B3LYP/6-311G** level of theory for the calculation of the spectroscopic constants of the protonated adducts of the OC_nO series (Cheikh and Pauzat 1999).

A last problem to solve in this case is to find the protonation site. Fixation of the proton is indeed possible at the terminal oxygen atom or at any carbon along the chain. It has been found that, whatever n , the most stable ion has the $\text{OC-CH-C}_{(n-2)}\text{O}^+$ structure. These compounds are singlet states, whatever the spectroscopic state (singlet or triplet) of the parent compound. The predicted dipole moments and spectroscopic constants are reported in Table 14.7.

14.2.1.3 Search for the P-C Bond in Space: HPCO and HPCS as Possible Tracers

If the formation of nuclear acids is a mandatory step on the path of emerging life, then, there must exist simple chemical systems that constitute the primary bricks of phosphorus compounds. A synthesis of these acids has recently been achieved in the laboratory (De Graaf et al. 1995), but the non-detection of polyatomic compounds with phosphorus is still considered a major difficulty for exobiology models. Only two molecules including phosphorus, PN and PC, have been detected in the interstellar medium (ISM). Neither hydrogenated compounds such as HPN, HPC, PNH or PCH (and their corresponding ions), nor PO or PS which are well-identified in the laboratory, have been observed. On the other hand, complex molecules with a P-C bond have been found in objects such as meteorites (for example in the Murchison meteorite (Cooper et al. 1992)). These are derivatives of phosphonic acid, HPO_3H_2 , whose hydrogen on the central phosphorus atom has been replaced by hydrocarbon groups; they might be the ancestors of the biological phosphonates. As of today, there is still a missing link in the chain leading to phosphorus prebiotic species.

By analogy to the [C,H,O,N] system, the simplest system that can possibly lead to the formation of the peptide bond, one may think of [C,H,O,P] as the simplest system leading to the carbon-phosphorus bond. Consideration of all the arrangements feasible with these four atoms, in triplet as well as in singlet electronic states, made it possible to show that singlet HPCO is the most stable structure from an energetic point of view.

The strategy employed to determine the rotational constants is based on high-level electronic calculations using coupled cluster or DFT methods, followed by a calibration procedure. Combining the experimental rotational constants B and C of HOCO^+ (Bogey et al. 1984, 1986) and HNCO (Hocking et al. 1975) with the values calculated for these two isovalent systems (same number of valence electrons), it is possible to associate a correcting factor $B_{\text{iso}}^{\text{Theory}}/B_{\text{iso}}^{\text{Exp}}$ to each level of theory. Such a calibration technique can only be validated by the convergence of the results, which provides at the same time an estimate of the error

Table 14.8. Rotational constants (GHz) for HNC0 and HPC0

Model	HNC0		HPC0	
	B	C	B	C
Ab initio				
CCSD/6-311G**	11.085	10.932	5.5165	5.4005
CCSD/6-311++G**	11.068	10.915	5.5212	5.4049
DFT				
B3LYP/6-311G**	11.091	10.954	5.5357	5.4169
CCSD/6-311++G**	11.078	10.943	5.5017	5.3842
Observed/Predicted	11.071	10.911	5.519	5.402

bar. The most significant results are reported in Table 14.8 for the lowest energy structures in the HNC0 and HPC0 series (Dimur et al. 2001).

These spectroscopic constants (with an error bar of $\sim 0.3\%$) should be precise enough to support the identification of the molecule in the laboratory.

14.2.2 Infrared Observations

A formidable challenge in modern astrophysics is the interpretation of the infrared emission of our galaxy between 2 and $15\mu\text{m}$ (Fig. 14.9). These bands, referred to as unidentified infrared bands (UIR) in the literature, were first discovered by ground-based observations at the end of the 1970s. Since then, their existence has been confirmed in a number of objects by observations using satellites IRAS and ISO (Cohen et al. 1986; First ISO Results 1996).

The same type of emission has been observed in various environments such as reflection nebulae, planetary and protoplanetary nebulae, H_{II} regions and galactic nuclei. This emission is characterized by a series of bands that are systematically found at 3.3, 6.2, 7.7, 8.6, 11.3 and $12.7\mu\text{m}$. Since IR frequencies are mainly specific of functional groups, the role of the numerical simulations is not to take the observed spectral signature and look for a hypothetical species that could be at its origin, but to determine the family of compounds that is responsible of the emission.

14.2.2.1 The PAH Hypothesis

The size distribution of interstellar particles is constrained by the black body thermal emission. The analysis of the emission between 2 and 15 microns can be interpreted as originating from very small grains composed of about ~ 50 atoms (hydrogens not taken into account) heated to temperatures on the order of ~ 1000 K after absorption of a photon. Larger species contribute to the spectrum at longer wavelengths.

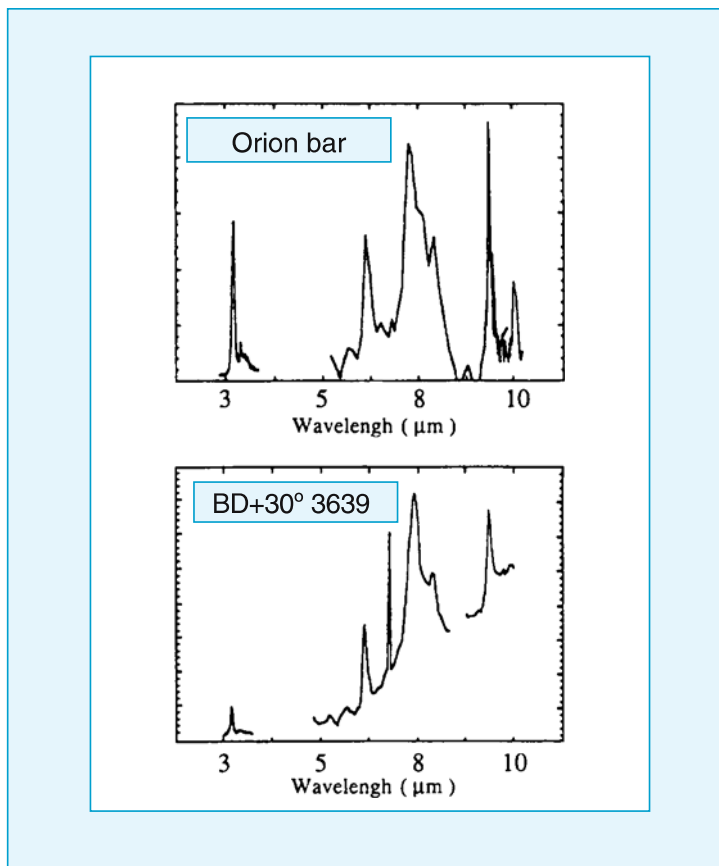


Fig. 14.9. Infrared emission from the Orion nebula and a planetary nebula (adapted from Schutte et al. 1993)

For interpreting this emission, Leger and Puget (1984) and Allamandola et al. (1985) proposed a model based on the results known in the IR spectroscopy of organic compounds. These features could be associated with small hydrocarbon clusters containing hydrogen atoms, more precisely to polycyclic aromatic hydrocarbons (PAH). These molecules could become vibrationally excited by absorption of UV photons and then release their excess energy by means of infrared fluorescence.

This hypothesis relies on a similarity between the spectra observed in space and those of PAHs obtained in the laboratory, whose structural assignments are given in Table 14.9. Contrary to radio spectra where there is a perfect match between observation and measurements that make identification possible, the detail of the spectral match in the infrared is not satisfactory, either for the band positions or for the relative intensities. Various interpretations of these dif-

Table 14.9. Assignment of the IR emission within the PAH hypothesis

$\nu(\text{cm}^{-1})$	$\lambda (\mu\text{m})$	Assignment	Symbol
3040	3.29	Aromatic C–H stretching	$\nu(\text{CH})$
1615	6.2	Aromatic C–C stretching	$\nu(\text{CC})$
1315–1250	7.6–8.0	Blending of aromatic C–C–C in plane deformations	$\nu(\text{CCC})$
1150	8.7	Aromatic C–H in plane bending	$\delta(\text{CH})$
885	11.3	Aromatic C–H out-of-plane bending (non-adjacent H)	$\delta(\text{CH})$
840	11.9	Aromatic C–H out-of-plane bending (2 adjacent H)	$\delta(\text{CH})$
790	12.7	Aromatic C–H out-of-plane bending (3 adjacent H)	$\delta(\text{CH})$
740	13.5	Out-of-plane C–C–C–C torsion	$\nu(\text{CCCC})$
740	13.5	Aromatic C–H out-of-plane bending (4 adjacent H)	$\delta(\text{CH})$

ferences have been proposed depending on the authors, leading to contradictory conclusions such as severe dehydrogenation or negligible dehydrogenation of the interstellar PAHs. It has been said that PAHs could be either neutral or positively charged due to the radiation field. It has also been suggested that they could be present in the form of negative ions within dense clouds where electrons seem abundant.

Given such a diversity of hypotheses about the structure of PAHs, the best strategy would be to measure the IR emission of a selected number of representative species and corresponding ions in the gas phase, and at temperatures and densities reproducing the interstellar conditions of the observed sources. It is a practically impossible task and numerical simulation is here an alternative.

14.2.2.2 Test with a Dehydrogenated Aromatic Species

The first point to be verified, as in any numerical simulation, is the adequacy of the method to the problem at hand. After a series of calculations on small standard PAHs (Fig. 14.10), it was important to test the first proposition made to interpret the weakness of the feature at $3.3\mu\text{m}$, i.e., the partial dehydrogenation of PAHs.

The simplest system that is really significant is dehydrogenated naphthalene (Fig. 14.11). The calculated IR frequencies and structural assignments are reported in Table 14.10 together with the results of matrix experiments.

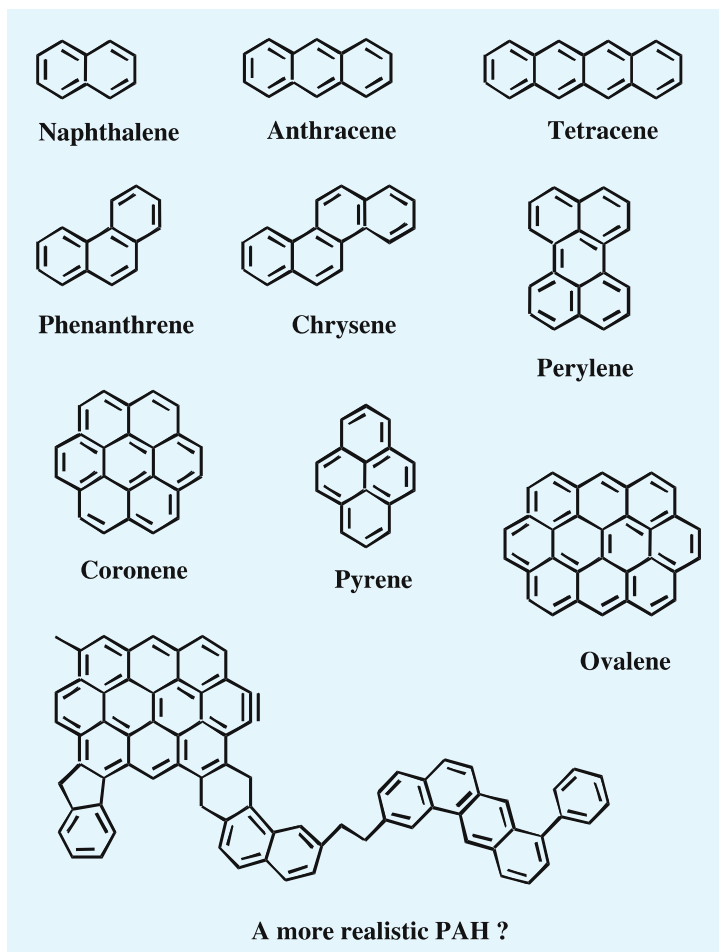


Fig. 14.10. Some PAHs of astrophysical interest

The results presented in Table 14.10 illustrate two independent studies, one theoretical (Pauzat et al. 1995) and one experimental (Weimer et al. 1995) that have been carried out independently. Correlation effects and anharmonicity corrections are not taken into account in the present calculations and a standard scaling procedure has been applied. The frequencies thus obtained, are close to those measured in the neon matrix experiment. The difference is only a few cm^{-1} for CH stretching vibrations; it is at most 3% for $\alpha(\text{CCC})$ in plane bending deformations that are the most sensitive vibrations to an incomplete treatment of electronic correlation. The intensities are only qualitatively reproduced, which is coherent with the level of theory used in the study. In spite of their shortcomings these calculations were able to confirm that a molecule, with a triple bond in

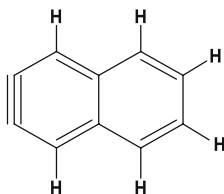


Fig. 14.11. Structure of 2,3-naphthylene

Table 14.10. Theoretical predictions and experimental values measured in a rare gas matrix. Absolute intensities being impossible to determine in the experiment, the most intense band of the spectrum has been arbitrarily given the corresponding theoretical value {92} for an easier comparison of relative intensities

Experiments in neon matrix		Theoretical predictions			
Frequencies ν (cm^{-1})	Relative Int. (km/mol)	Frequencies ν (cm^{-1})	$\Delta\nu\%$	Absolute Int. (km/mol)	Assignments
3098	8	3090	0.3	13	r(CH)
3084	10	3078	0.2	19	r(CH)
3069	16	3065	0.2	21	r(CH)
1213.1	4	1232	1.6	23	$\beta(\text{CH})/\text{R}(\text{CC})$
1018.3	2	1010	0.7	17	$\text{R}(\text{CC})/\beta(\text{CH})$
848.6	9	869	2.3	24	$\alpha(\text{CCC})$
834.1	37	843	1.0	74	$\varepsilon(\text{CH})$
738.6	47	739	0.0	51	$\tau(\text{CCCC})$
618.1	79	600	2.9	54	$\alpha(\text{CCC})$
451.8	9	446	1.3	22	$\tau(\text{CCCC})$
446.5	{92}	438	1.8	92	$\alpha(\text{CCC})$

an aromatic six-membered ring, 2,3-naphthylene, was effectively formed by irradiation of naphthalenic compounds in the rare gas matrix. The fact that the IR signature of such an exotic species is so well-reproduced by quantum chemistry simulations lends confidence in forthcoming applications to other molecules in space.

14.2.2.3 Prediction of the IR Signature of PAHS as a Function of Ionization

As mentioned previously, it has been proposed, among other ideas, that PAHS could exist, not only as neutral entities but also as positive ions, as a reason

of the radiation field to which they are submitted (Allamandola et al. 1985; d'Hendecourt and Léger 1987). Formation of doubly-charged PAHs in space by successive ionizations was first considered by Leach (1986). Laboratory experiments have shown that the second ionization potential of PAHs is generally lower than 13.6 eV, i.e., lower than the limit of ionization for atomic hydrogen (Tobita et al. 1994). The abundance of PAH di-cations in the diffuse interstellar medium has not yet been evaluated because they are the outcome of a competition between the second ionization and the recombination with the electrons, which is a process difficult to estimate.

Theoretical calculations including correlation effects at the DFT/B3LYP level of theory have been performed for neutral pyrene Py and the corresponding cation Py^+ and di-cation Py^{++} (Ellinger et al. 1999). The simulated spectra are presented in Fig. 14.12. They show unexpected features for the ionized systems. It can be seen that the frequencies are little affected by ejection of one or two electrons (shifts in frequency $< 50\text{cm}^{-1}$), but there are extraordinary variations in the intensities. The most spectacular point is the collapse of the intensity of the feature associated with the CH stretching vibration at $3.3\mu\text{m}$ following the first ionization.

This result is well in line with the fact that, as of today, there is no information available on the $3.3\mu\text{m}$ region in the case of the PAH cations. Ejection of two electrons leads to an intensity of the CH stretching that is divided by a factor of 2 or 3 with respect to the neutral molecule, which means that the intensity increases with respect to the singly-ionized PAHs.

The origin of the intensity collapse of the CH stretching vibration can be rationalized if one considers that the vibration is sufficiently localized to be represented by a diatomic model. At the same DFT/B3LYP level of theory, the IR intensities are very different for CH (185kmmol^{-1}) and CH^+ (1kmol^{-1}). It is directly related to the ionization of the electron in the $2p$ orbital of the carbon in the CH fragment, which is equivalent to creating a hole in the π system of aromatic compounds. The effect of the double ionization cannot be modeled by a CH fragment and remains without an explanation.

Comparison of absolute intensities (which cannot be done in Fig. 14.12 since each spectrum is normalized to its own most intense band) shows a very strong increase in the intensities of the $\beta(\text{CH})$ bands about 1240cm^{-1} following the first ionization: 3kmmol^{-1} (Py); 78kmmol^{-1} (Py^+); 3kmmol^{-1} (Py^{++}). More generally, all in-plane modes of vibration have their intensities increase with the first ionization.

An interesting result has been found for out-of-plane CH vibrations. They are shifted by ionization but the intensities are not affected. For example, the $\varepsilon(\text{CH})$ band at (843cm^{-1} ; 97kmmol^{-1}) in neutral pyrene is found at (856cm^{-1} ; 97kmmol^{-1}) in the cation and at (871cm^{-1} ; 94kmmol^{-1}) in the di-cation. These values are to be compared with the experimental results of (843cm^{-1} ; 155kmmol^{-1}) and (861cm^{-1} ; 66kmmol^{-1}) for neutral and singly-ionized pyrene, respectively, in the neon matrix.

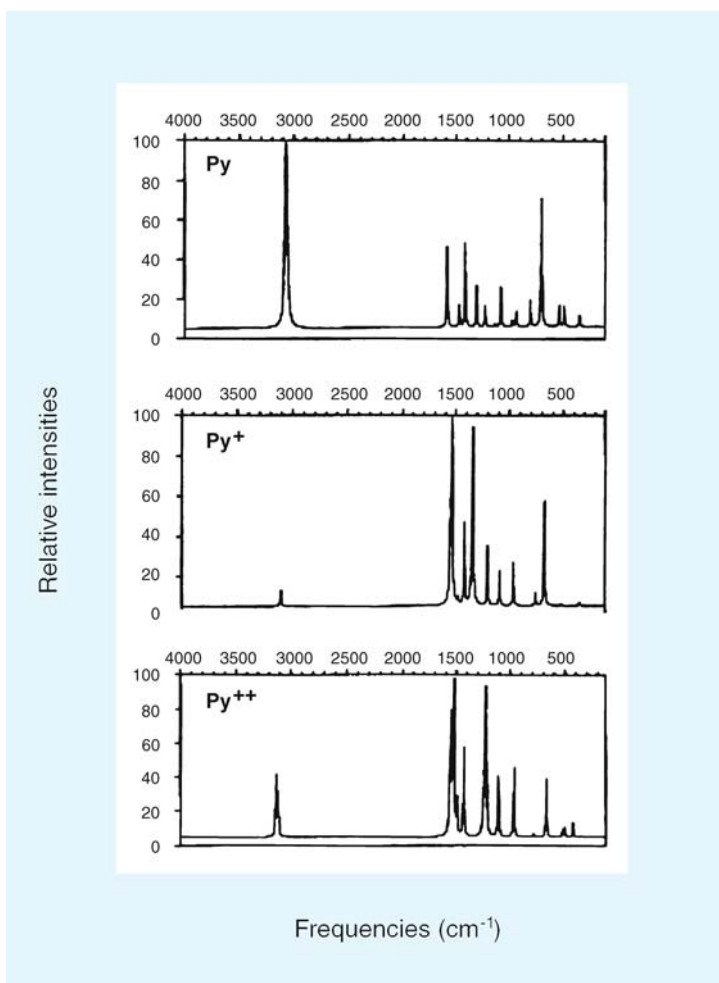


Fig. 14.12. Simulated spectra for neutral pyrene, and corresponding cation and dication

To summarize, the results of the quantum chemistry calculations are in excellent agreement with the experimental data presently available on pyrene and its positive ion. The difficulties encountered in the models to interpret the intensity ratios between the weak band at 3.3 μm and the very strong ones at 6–8 μm can be removed if one supposes that PAHs are partially ionized. It is a reasonable assumption according to the radiation field in the regions where they are observed.

The theoretical results presented here are part of more general studies of the effects of ionization on the IR spectra of PAHs (Pauzat et al. 1992, 1995; DeFrees et al. 1993; Vala et al. 1994; Langhoff 1996), the first theoretical studies being

done before and simulating the experiments on PAH⁺ positive ions (Szczepanski et al. 1992, 1993; Szczepanski and Vala 1993; Hudgins et al. 1994, 1995a,b; Hudgins and Sandford 1998). Besides, it should be added that the predictions made for doubly-ionized species are of a quality at least equal to that already achieved for singly-ionized species for the simple reason that, contrary to singly-ionized PAHs that are open-shell systems, di-cations are closed-shell systems that do not suffer from spin contamination.

14.2.2.4 Effects of Substitution on the IR Spectra of PAHs

The region around 3.3 μm is certainly the one that has been the object of most controversy. This feature was first observed as a single peak of variable intensity according to the IR sources. With the increase in the precision of the instruments, a second component was discovered at 3.4 μm , that appears to be the dominant peak over a broad plateau spread over the 3.35 to 3.6 μm region, with numerous bands now well resolved.

It is clear that the PAH hypothesis must be refined if a complete account of the observations is to be made in this context. Several interpretations have been proposed:

- The first interpretation is based on the so-called hot bands hypothesis (Barker et al. 1987). In this case, the series of observed bands not only comes from the transition ($v = 1 \rightarrow 0$) but implies transitions between more excited levels ($v = 2 \rightarrow 1$; $v = 3 \rightarrow 2$, etc.). It is the anharmonicity of the CH stretching potential that would be at the origin of the shifts observed. The existence of a peak about 1.68 μm does correspond to the laboratory spectrum of the ($v = 2 \rightarrow 0$) transition. It is strong support for the hypothesis, but the relative intensities at 1.68 and 3.34 μm show that the hot bands alone cannot account for the intensity of the 3.34 μm feature. Furthermore, it is not clear that what has been seen for neutral species can be extended to ionic ones.
- The second interpretation proposed considers the possibility of specific vibrations of chemical groups attached to the periphery of the PAHs, (Duley and Williams 1981). The first idea was that of a CH vibration from an aldehyde group HC = O. It has been rejected due to the non-observation of the associated C = O stretching vibration. By contrast, the hypothesis of alkyl substituents appears more justified if one considers the results of the experiments performed on methyl and ethyl coronene. Close to the peak at 3.3 μm there are similar satellite bands extending between 3.35 and 3.55 μm (Joblin et al. 1996).
- The third interpretation is that of the hydrogenated PAHs (Schutte et al. 1993). The addition of hydrogen atoms to PAHs results in a rupture of the π system of the aromatic rings. Tetrahedral centers are then created with one or two hydrogens according to the position of the carbons. It has been shown that the stretching vibrations of such hydrogens are closer to aliphatic

hydrogens than to aromatic ones and that the corresponding frequencies also cover the 3.35 and 3.55 μm plateau.

Numerical simulation is a powerful tool that is well-suited to the study of frequency shifts, whatever the family of compounds considered, neutral or ionized PAHs. Only the second interpretation will be considered here as an example of this type of research.

Normal coordinate analysis of vibrational spectra makes it possible to assign the observed frequencies (and intensities) to precise chemical groups in the case of substituted PAHs (aromatic CH, $-\text{CH}_3$, $-\text{CH}_2\text{CH}_3$, $-\text{CH}=\text{CH}_2$). Although a definite assignment to a particular molecule cannot be done using only the calculated values, it is nevertheless possible to show what type of substituted PAH is at the origin of a given peak. An example of this theoretical approach is reported on Fig. 14.13 giving an illustration of the effect of alkyl side groups on the spectra in the 3.35 and 3.55 μm region (Pauzat Ellinger 2001).

The important conclusion that can be drawn from the simulations is that PAHs with CH_3 groups attached at the periphery can certainly be at the origin of the satellite bands at $\sim 3.4 \mu\text{m}$. Both calculations and laboratory experiments place these bands in the 3.35–3.45 μm region. In addition to a correct position, the calculations provide another strong argument in favor of the side group hypothesis: the strong intensity of the CH stretching vibrations of the CH_3 and C_2H_5 groups compared with the CH aromatic stretching vibrations. It is important for neutral PAHs and even more for ionized ones since the collapse of the intensities is less pronounced for the substituents than for the ring hydrogens. In this context, a small quantity of substituted PAHs is sufficient to generate a significant emission around 3.4 μm .

The results of the ab initio numerical simulation are in excellent agreement, for both frequencies and intensities, with the experimental spectra of anthracene and pyrene derivatives taken as model PAHs in their neutral and ionized forms to account for the different radiation fields. This study shows that the variety of spectra observed in the region at 3.3 μm can effectively be interpreted within the PAH hypothesis when considering simple aromatic species and their substituted derivatives. The conclusions can be summarized as follows:

- The peak at 3.4 μm can be associated with the $-\text{CH}_3$ group.
- The peak at 3.5 μm can be associated with combination bands implying the vibrations of angle deformation of the $-\text{CH}_3$ umbrella but the weakness of the intensity suggests the contribution of other species.
- The peak at 3.2 μm due to $=\text{CH}_2$ groups has a weak intensity that makes the observation difficult; the fact that it has not been observed indicates that there should be few of these side groups, in agreement with photochemical models.
- The intensity of the satellite bands is not significantly affected by the ionization process contrary to the 3.3 μm aromatic CH vibration.

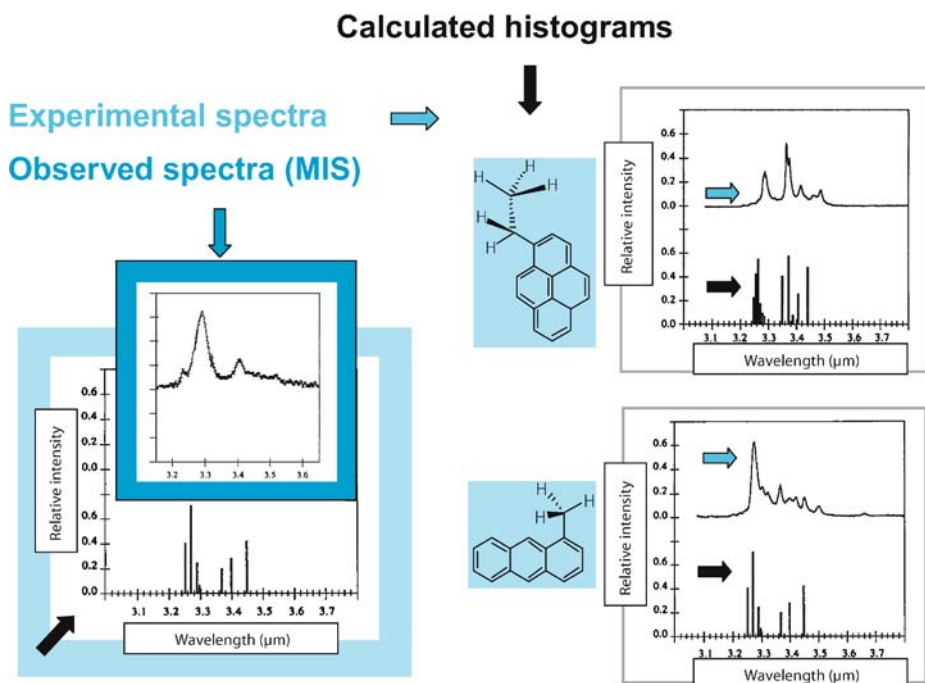


Fig. 14.13. Gas phase spectra of methyl anthracene and ethyl pyrene in the 3.2–3.6 μm region compared to observations of NGC 1333 and histograms obtained by ab initio numerical simulation (adapted from Joblin et al. 1996; Pauzat and Ellinger 2001)

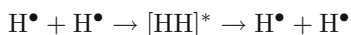
This section illustrates the obvious advantages of the numerical simulation. Regarding the chemical modeling, it makes it possible to determine a priori the spectral signatures of molecules that are too unstable to survive in the laboratory conditions such as free radicals or dehydrogenated structures. The same strategy of numerical simulation can be applied whatever the level of ionization, i.e., for neutral, singly or doubly-ionized species as well as for negative ions. Regarding the structural modeling, the normal coordinates analysis makes it possible to assign the observed frequencies to precise vibrations, and determine the structural origin of the observed peaks.

14.2.3 Modeling of Chemical Processes

The first step in understanding the chemical evolution of molecular compounds toward the formation of the prebiotic bricks at the origin of life is to detect and identify the molecules actually present in the interstellar medium. However, interstellar chemistry that takes place in extreme conditions of temperature and pressure requires an original way of thinking, quite unusual to laboratory chemists. The first thing to realize is that there is no energy thermostat in the

ISM gas phase: nothing in the environment can furnish or dissipate the energy. It follows that the energy released in an exothermic reaction remains a priori confined inside the system; it can only be dissipated by radiation emission (generally a slow process) or by breaking the system that has just been formed so that the excess energy can be eliminated in the form of the kinetic energy of the fragments.

The most spectacular example is that of the synthesis of molecular hydrogen. The collision of two hydrogen atoms



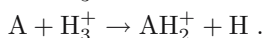
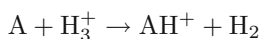
leads to a diatomic complex with an internal energy of ~ 104 kcal/mol corresponding to the formation of the H–H bond. Because this energy cannot be evacuated, the complex cannot stabilize and H_2 dissociates giving back two hydrogen atoms. The presence of a third body to absorb the energy released is a compulsory requirement for the formation of the most abundant molecule of the universe. This example makes clear the importance of dust particles or interstellar grains that are thermodynamically necessary to favor chemical processes at the solid/gas interface that would not be feasible otherwise in the gas phase.

About 140 molecules have been identified in space. They can be classified into two categories: neutral and ionic species. Two kinds of elementary processes can therefore be considered depending on whether they involve compounds of one or the other type.

14.2.3.1 Modeling of Reactions Involving Ions in the Gas Phase

Ionized hydrogen is certainly the most abundant positive ion in the universe, giving its name (H_{II} regions) to environments submitted to a strong radiation pressure. In less exposed regions, the H^+ ion is associated with molecular hydrogen in the form of H_3^+ .

It is mainly under this form that the proton reacts since it is possible to evacuate the excess energy released in the proton transfer reaction as kinetic energy of the H_2 fragment or a hydrogen atom after fixation of H_2^+ . The corresponding chemical equations are



The network of reactions reported on Fig. 14.14 is an illustration of a chemical model based on a series of reactions between neutral molecules, mainly H_2 and N_2 , and positive ions. The two positive ions at the origin of the modeling are H_3^+ and N^+ , the latter coming from the ionization of the nitrogen atom by cosmic rays (RC) or from the destruction of N_2 by He^+ .

In this model, the synthesis of complex systems progresses step by step, the excess energy being evacuated by expulsion of a hydrogen atom in both cases of reactions with H_2 or electronic dissociative recombinations.

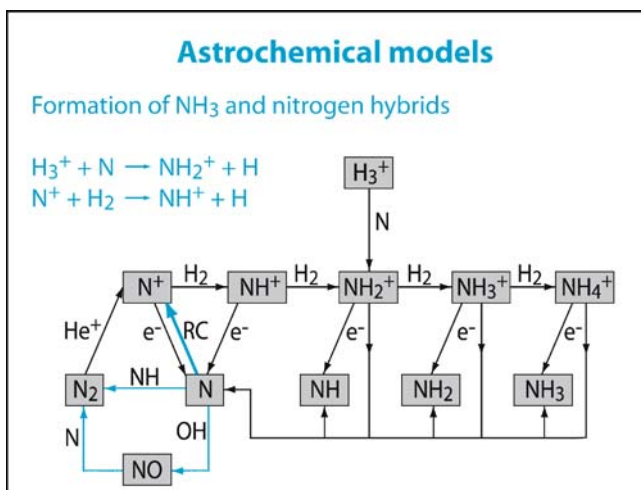
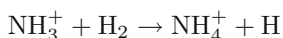


Fig. 14.14. Modeling of the formation of ammonia and nitrogen hydrides

a) Variation of a Rate Constant with Temperature

It is generally assumed that the rate constant of an ion-molecule reaction increases with temperature. This temperature dependence, commonly applied in chemical models, is not as universal as usually thought. For example, the reaction



whose energy profile is given in Fig. 14.15, exhibits an unusual dependence with temperature (Luine and Dunn 1985). The rate constant k does not follow the Arrhenius dependence for temperatures below 300K: it is more or less even in the 80–100K interval (Fig. 14.16) and then increases unexpectedly at lower temperatures. In brief, one has

$$\begin{aligned}
 10 \text{ K} < T < 100 \text{ K}, & \quad k \text{ decreasing} \\
 100 \text{ K} < T < 300 \text{ K}, & \quad k \text{ increasing.}
 \end{aligned}$$

The theoretical scheme that has been proposed to explain this unusual phenomenon consists of the primary formation of a complex on the entrance channel that, at low temperatures, could transform into the reaction products by means of tunneling effects.⁶

Numerical simulation methods have the capability of probing such ideas. The potential energy surface has been calculated for a molecule of H₂ approaching the NH₃⁺ ion. The energy profile reported on Fig. 14.15 shows an activation

⁶ In a reaction proceeding via a tunneling effect, there is a non-zero probability for the reaction path to go *through* the potential barrier. It is not necessary to have enough activation energy to go *over* the barrier; it is a quantum effect.

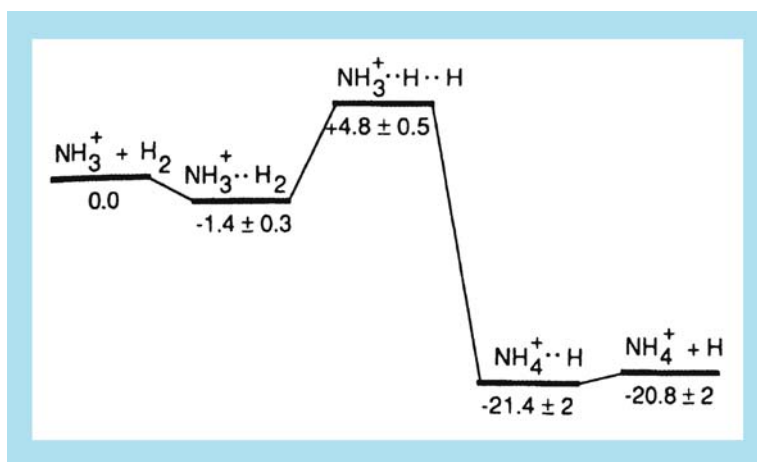


Fig. 14.15. Minimum energy profile for the reaction $\text{NH}_3^+ + \text{H}_2 \rightarrow \text{NH}_4^+ + \text{H}$ (energies in kcal mol⁻¹)

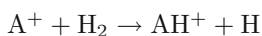
barrier slightly inferior to 5 kcal mol⁻¹ with two weakly bound complexes, one on each side.

In order to cover all possibilities of approach, the orientation of H_2 has not been submitted to any constraints. The postulated complex between the neutral molecule is effectively found on the entrance channel with a stabilization energy of ~ 1.4 kcal mol⁻¹. The second complex on the exit channel is more weakly bound. The exothermicity of the reaction is ~ 21 kcal mol⁻¹. At very low temperatures, the lifetime of the $[\text{NH}_3^+ \cdots \text{H}_2]$ complex is non-negligible. The lower the temperature, the longer the lifetime, which leaves a longer time for the tunneling to be effective. As the temperature increases, the energy becomes greater than the activation barrier and the rate constant recovers the Arrhenius dependence. It is to be noted that the exit channel complex is not stable enough to influence the reaction rate.

Such a potential surface, that links a minimum energy structure on the entrance channel to the products through a well-defined transition state (one imaginary frequency) is suitable for applying RRKM theory to the calculation of the rate constant. Tunneling through the barrier (described by an Eckart potential) is also treated using an extension of RRKM theory.

The numerical simulation (Herbst et al. 1991) shows that the reaction is controlled by a competition between two mechanisms. On one side, one has the back dissociation of the first complex toward the starting products, which depends of the lifetime of the complex; on the other side, one has the evolution toward the products through (tunneling) and over the barrier (classical process).

These results suggest that other ion-molecule reactions of the type



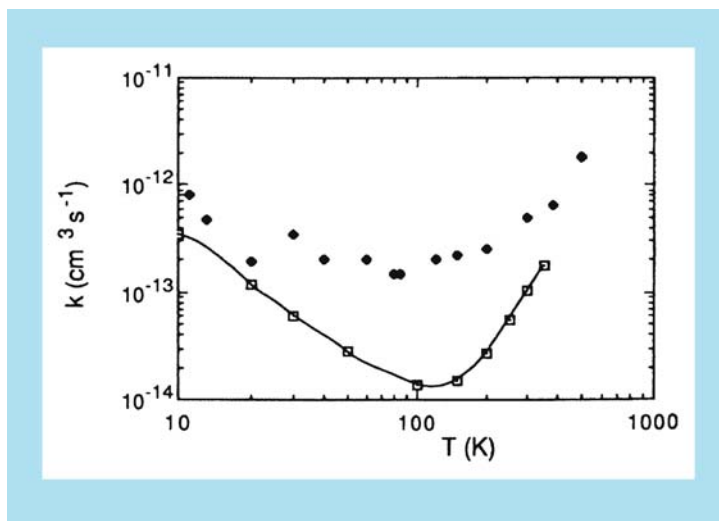


Fig. 14.16. Evolution of the rate constant k for reaction $\text{NH}_3^+ + \text{H}_2 \rightarrow \text{NH}_4^+ + \text{H}$ as a function of temperature. The solid curve corresponds to the theoretical values. Experimental points are represented by black diamonds (the experimental uncertainty is on the order of 50%)

could well present the same inverse dependence with T at very low temperatures (Fig. 14.16). If this phenomenon were more general, then one can anticipate a strong impact on the chemical models involving ion-molecule reactions in the interstellar medium in which molecular clouds are at the same time subject to very low temperatures (10–70 K) and dominated by the abundance of molecular hydrogen.

b) Reactions on Potential Surfaces of Different Multiplicities

Atomic nitrogen, once ionized by cosmic rays, does not react only with hydrogen. In most models of interstellar chemistry, this atomic ion is at the origin of the formation of molecules containing the CN bond (Talbi 1999). In these reactions, the partner is the CH_3 free radical. In such a case, one expects a very fast reaction without an activation barrier. The obvious reason is that the reaction involves a positive ion and a very reactive species, i.e., a free radical with an unpaired electron (Fig. 14.17).

Another advantage of numerical simulations is the intrinsic capability of following how reactions proceed on potential energy surfaces associated with different electronic states and/or multiplicities. The reaction profiles reported on Fig. 14.17 illustrate a case of doublet and quadruplet multiplicities.

On the doublet surface, the first step is the formation of the HCNH_2^+ molecular ion that rearranges into $\text{HCNH}^+ + \text{H}$ by going over an activation barrier of 29 kcal mol^{-1} . This barrier is easily passed due to the $229 \text{ kcal mol}^{-1}$ released in

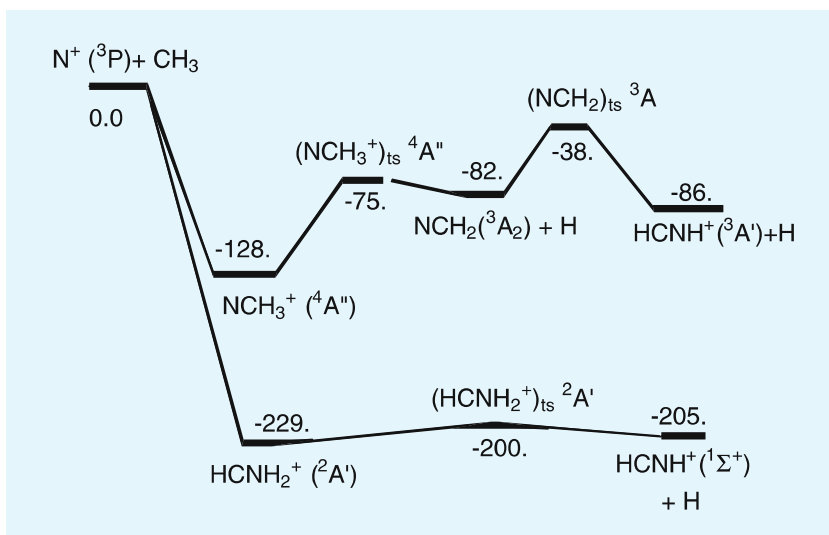


Fig. 14.17. Energy profile for the reaction of N^+ with CH_3 (energies in kcal mol^{-1}). The *ts* index refers to transition states

the first step of the reaction. On the quadruplet surface, there is first an addition of N^+ to CH_3 , followed by the ejection of a hydrogen atom to give NCH_2^+ . This ion may then undergo an internal rearrangement leading to the final product, ${}^3\text{HCNH}^+$. It should be noted that this second reaction path leads to an excited triplet state that is, at the same time, higher in energy and subject to a higher activation barrier (90kcal mol^{-1}). Although the two surfaces lead to the same chemical species, they should be considered separately: one results in the direct formation of the ${}^1\Sigma^+$ ground state (doublet surface), the other gives an excited ${}^3A'$ state (quadruplet surface) whose electronic de-excitation toward the ground state must be considered as an independent process. Since the favored reaction path is that of lower energy, the final product is HCNH^+ (the H_2CN^+ ion is not formed in the reaction). Other ion molecule reactions also lead to HCNH^+ , such as $\text{C}^+ + \text{NH}_3$ where ammonia may be formed by the chain of reactions given as an example on Fig. 14.14.

c) Reactions of Electronic Dissociative Recombination of HCNH^+

The neutralization of positive ions by addition of one electron is the major link between ions and neutral species. In the case of the HCNH^+ ion, two possibilities must be considered, one leading to HCN , a well-known molecule on Earth, the other leading to the HNC isomer, unstable on Earth but quite abundant in a number of astrophysical objects.

This process implies the capture of an electron by the HCNH^+ ion that results in the formation of the corresponding neutral system in an excited state. If this

state is repulsive, there is a direct dissociation with expulsion of a hydrogen atom. If not, two possibilities should be considered: either an autoionization, or a relaxation toward lower energy states (mainly Rydberg states⁷) followed by fragmentation. This process is said to be indirect.

In the numerical simulation, the main difficulty resides in the obligation to treat the ion and the excited neutral system in its stable Rydberg and dissociative valence states with an equivalent precision (Talbi et al. 1989). The simulation must be performed for the two possible dissociation paths: the rupture of the N–H bond and the rupture of the C–H bond. The method is based on the construction of a wave function that is able to describe the system in terms of localized orbitals representing the chemical formula (Fig. 14.18).

By using a localization procedure, it is possible to generate a set of bond orbitals σ_{CH} , σ_{NH} , σ_{NC} , π_{NC} , π_{NC} satisfying the chemical intuition without loss of generality together with the necessary Rydberg type orbitals to build the excited states; these are taken in the form of diffuse functions centered in the middle of the C \equiv N bond and spreading out far from the molecular backbone. As shown previously in the theoretical part, one has to generate the associated antibonding orbitals σ_{CH}^* , σ_{NH}^* , σ_{NC}^* , π_{NC}^* , π_{NC}^* to properly account for correlation effects on the potential surfaces describing the rupture of the C–H and N–H bonds. In these calculations, the wave function is built so as to treat all excited configurations, the Rydberg and the dissociative, on an equal footing.

The most important correlation effects are taken into account by considering the mono-di-tri-quadi excitations of the electrons from the orbital space occupied in the neutral molecule to the virtual space composed of the antibonding and Rydberg orbitals.

The expansion of the multiconfigurational wave function in terms of localized orbitals representing the chemical formula allows, for every point of the adiabatic

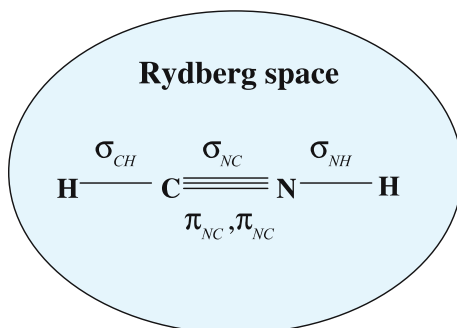


Fig. 14.18. The valence orbitals used to describe the process of dissociative recombination in a Rydberg space

⁷ Electronic state in which the excited electron is sufficiently distant from the molecular skeleton to feel the attraction of the positive core as if it were an atomic cation.

surfaces, the assignment of each potential surface to a chemically well-defined state. Only the surfaces that are important in the process of dissociative recombination have been represented on Fig. 14.19. A careful examination of these surfaces shows that the curves of the dissociative states cross those of the ion at the minimum of the CH elongation; the direct mechanism will then be preponderant and will take place in an equivalent way for the two reaction paths. The same situation prevails for the indirect mechanism since the dissociative curves cross the Rydberg states at the same level (Talbi and Ellinger 1998).

The conclusion of this study is that the electronic dissociative recombination of the HCNH^+ ion gives two neutral species, HCN and HNC, in equivalent quantities in spite of their very different stabilities. It is one paradox of interstellar chemistry.

14.2.3.2 Modeling Gas Phase Reactions Implying Neutral Molecules

For a long time, models of interstellar chemistry have neglected neutral-neutral reactions that were supposed to be slow compared to ion-molecule processes. Another reason is the lack of experimental rate constants due to the difficulty of handling non-charged species in the laboratory. From a theoretical point of view, these reactions can be calculated as easily as (or with no more difficulties than) the reactions involving ions.

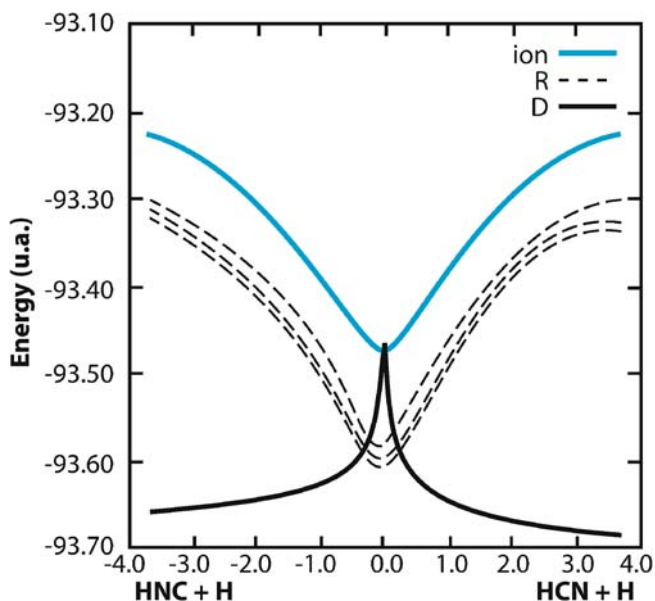
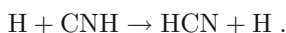


Fig. 14.19. Quasidiabatic representation of Rydberg and dissociative state of HCNH^+ ; *R*: Rydberg state; *D*: dissociative state

In the line of the preceding section, it is interesting to consider what happens to the two isomers HCN and HNC that have been formed in equivalent quantities. Observations in two different objects show intriguing results:

1. In TMC-1, an example of a dark cloud, HNC is dominant with an abundance ratio $\text{HNC}/\text{HCN} = 1.55$, which is surprising in light of the difference in the stabilities of the two isomers.
2. In OMC-1, a star-forming region, HCN is largely dominant with an abundance ratio HCN/HNC about ~ 80 in the vicinity of Orion-KL, a result conforming more to the intuition of the chemist on Earth.

It has been proposed recently that reaction with atomic hydrogen, certainly the most abundant reactive species in the universe with the free electron, is at the origin of the equilibration of the abundances:



This process would not be effective at very low temperatures because of the presence of an activation barrier, but its contribution would become significant at higher temperatures. Here also, numerical simulation appears to be an alternative to the difficult experiments on HCN (Talbi et al. 1996).

As before, the strategy is to calculate the approach of the hydrogen atom on the carbon side of the CNH molecule and follow the energy variation of the system until the expulsion of the hydrogen initially attached to the nitrogen atom. This push-pull or addition-elimination mechanism has an energy profile that satisfies the expectations (Fig. 14.20).

The two approaches, on the carbon side and nitrogen side, lead directly to transition states, without formation of long distance complexes. The best estimates of the activation barriers are 4.2kcal mol^{-1} on the $\text{H} + \text{CNH}$ entrance channel and 18kcal mol^{-1} on the $\text{H} + \text{NCH}$ entrance channel. The transition state $(\text{H-CN H})_{\text{TS}}$ is $\sim 2\text{kcal mol}^{-1}$ higher in energy than the $(\text{HCN-H})_{\text{TS}}$.

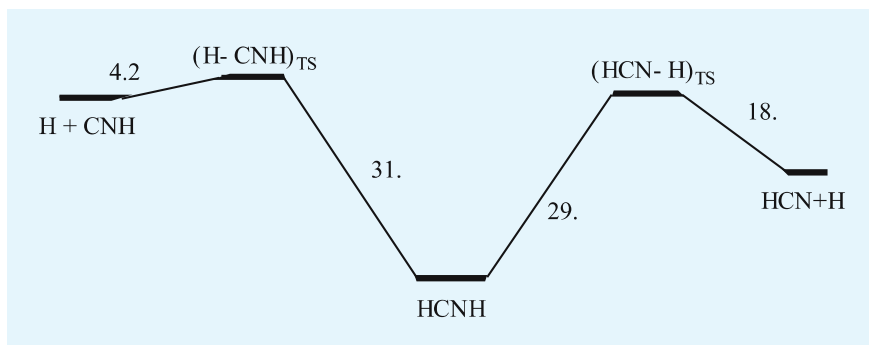


Fig. 14.20. Energy profile for the $\text{H} + \text{CNH} \leftrightarrow \text{HCN} + \text{H}$ hydrogen exchange reaction (energies in kcal mol^{-1}). The TS index refers to transition states

These results have two important consequences:

1. The fact that the activation energy for the $\text{HNC} + \text{H}$ reaction is lower than that of $\text{HCN} + \text{H}$ means that there is a range of temperatures for which the direct process will transform HNC while the inverse process that transforms HCN cannot operate.
2. The fact that the transition states have different energies, has the consequence that, once the energy saddle point $(\text{H-CN})_{\text{TS}}$ is passed, there is no more obstruction to dissociation in $\text{HCN} + \text{H}$. The formation of the intermediate HCNH structure needs the presence of a third body (or an internal relaxation that should be marginal in the present conditions). By contrast, the inverse process has two activation barriers opposing the destruction of HCN to rebuild CNH.

The rate constant of this exchange reaction has been calculated using an RRKM method coupled with a perturbative treatment of tunnel effects for temperatures below 300K. The results of the calculation are presented on Fig. 14.21.

For temperatures below 300K the variation of the rate constant is similar to that already obtained for $\text{NH}_3^+ + \text{H}_2$. The reaction rate decreases with temperature until $\sim 50\text{K}$ and starts increasing gently for lower temperatures.

For temperatures above 300K the rate constant increases in a standard way with T to reach typical values of $(10^{-10} - 10^{-11} \text{ cm}^3 \text{ s}^{-1})$ around 700K. This

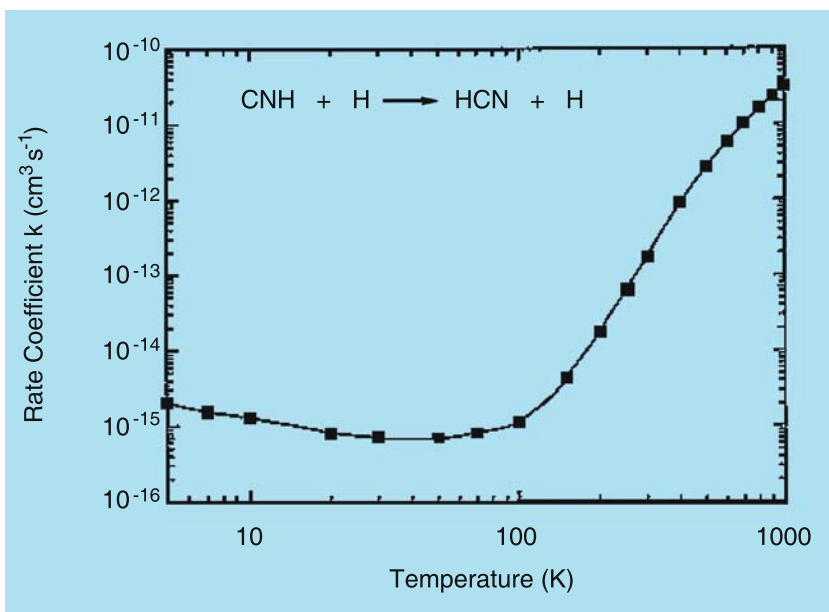


Fig. 14.21. Variation of the rate constant for the $\text{H} + \text{CNH} \rightarrow \text{HCN} + \text{H}$ reaction as a function of temperature

variation with the temperature is much more complex than postulated in the first generation of chemistry models and only the numerical simulation made it possible to identify this unexpected behavior.

It should be stressed, however, that this mechanism is not sufficient to account for the variation of the abundance ratio HCN/HNC, even if one takes an error bar on the order of 1 kcal mol^{-1} for the activation barriers. The $\text{H} + \text{CNH}$ reaction cannot explain alone the large value of the abundance ratio HCN/HNC observed, unless the temperature is much higher than generally assumed in regions considered. One has to find other types of reactions to rationalize the small quantities of CNH, for example investigate the role of interstellar grains and ices.

14.2.4 Exobiology

Since the first modern theory developed by Haldane (1929) on the subject, the interest in the first steps of emerging life has attracted increasing attention. Experiments were devised in the 1950s by Miller's group (Miller 1953; Miller and Urey 1959) and Abelson (1959) to produce amino acids from various mixtures of simple molecules. In these experiments, the energy necessary to initiate the reactions was provided in a *violent form*, as electric discharges for lightning, particle beams for cosmic rays or projectiles simulating volcano eruptions. The common point is that in these conditions, there is enough energy available to reach the most stable species that can be formed.

About the same time, a precursor of the peptide bond, HNCO, was also identified in the interstellar medium (Snyder et al. 1972; Brown 1981) and more recently in comets Hyakutake and Hale-Bopp (Lis et al. 1997; Bockelée et al. 2000). The next step on the amino acids and protein route, $\text{H}_2\text{N}-\text{CH}=\text{O}$, i.e., the simplest possible species containing the peptide bond, was identified in space (Rubin et al. 1971) and in the comets mentioned above, but none of its isomers have been detected to our knowledge.

14.2.4.1 Formation of the Peptide Bond

The problem here is not the calculation of the spectral signatures of simple prebiotic species (they are well established in the laboratory), but to understand why they are not observed.

The simplest peptide bond is found in formamide, $\text{H}_2\text{N}-\text{HC}=\text{O}$. In the violent conditions of Miller experiments, all the isomers that can be formed with the six atom system $[\text{C}, 3\text{H}, \text{O}, \text{N}]$ are a priori possible. Five of them have a linear backbone, the last one is a cyclic molecule. Theoretical calculations are able to give the geometries and energies, which makes it possible to compare their relative stabilities (Fig. 14.22).

The most stable system, without any doubt, is formamide. The next isomer on the energy scale, $\text{HN}=\text{CH}-\text{OH}$, has the same $\text{N}-\text{C}-\text{O}$ connectivity; it is the corresponding enolic tautomer, located $\sim 11 \text{ kcal mol}^{-1}$ higher in energy. The

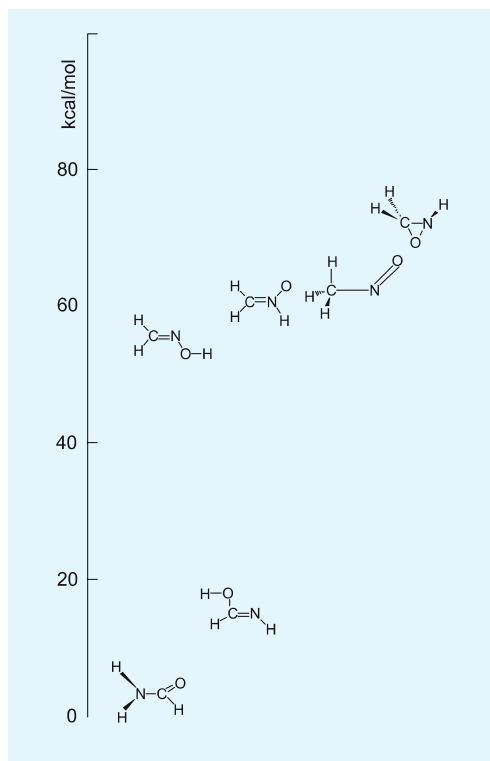


Fig. 14.22. Relative energies of the possible compounds of general formula $[C, 3H, O, N]$ with respect to formamide

third isomer, $H_2C=N-OH$ is formaldoxime; it is the first structure with a $C=N$ double bond. The next two isomers are nitrogen oxides: $H_2C=NH-O$, and $H_3C-N=O$, i.e., the nitron and the corresponding nitroso compound. The last isomer is a three-membered ring.

It should be noticed that the two isomers with $C-N-O$ connectivity, that is the atomic sequence of prebiotic structures, are by far the most stable. The other connectivities are more than 50kcal mol^{-1} above formamide.

It is also interesting to consider the case of the corresponding sulfur molecules of general formula $[C, 3H, S, N]$. As previously, the same six isomers are possible, namely, five open-chain structures and one cyclic molecule. The relative energies are reported in Fig. 14.23. It can be seen that thioformamide, $H_2N-HC=S$ and the corresponding enol $HN=HC-SH$ are the most stable structures. The differences on the energy scale are smaller, the thio-oxime structure being only 30kcal mol^{-1} above the most stable isomer.

This study (Chiaramello et al. 2005) demonstrates without ambiguity that the peptide bond that is found in biological systems effectively leads to the

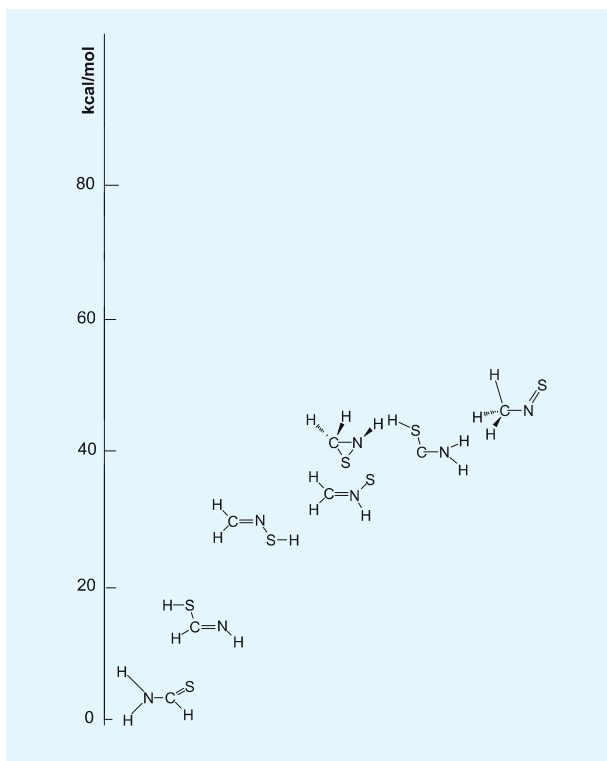


Fig. 14.23. Relative energies of the possible compounds of general formula $[C, 3H, S, N]$ with respect to thioformamide

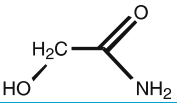
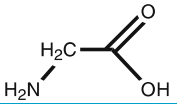
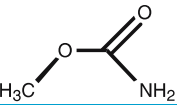
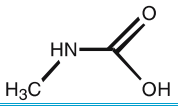
molecular structure of lower energy for both oxygen and sulfur-containing compounds. It is not by accident that these prebiotic systems are formed in extreme conditions: it is just because they are the most stable that can be formed.

14.2.4.2 Relative Stability of Amino Acids

The next step on the prebiotic roadmap is the formation of amino acids. The meteorites collected on Earth are an important source of extraterrestrial organic matter. A great number of amino acids have been found in the carbonaceous chondrites, including the 20 molecules present in living organisms. By contrast, glycine, $H_2N-CH_2-C(OH)=O$, which is the simplest of amino acids has not yet been discovered in space. The question to be settled is therefore that of the relative stability of the amino acid structure with respect to the other possible isomers.

The calculations performed on the few isomers that one might think of as playing a role in prebiotic chemistry provided the results given in Table 14.11.

Table 14.11. Relative energies of some of the possible isomers of glycine (energies in kcal mol⁻¹)

Name	Structure	Energy
Hydroxy-1 acetamide		13.6
Glycine		10.8
Methyl carbamate		4.2
N-methyl carbamate		0.0

The most stable isomer has both amine and acid functional capabilities, but it is not glycine. The next molecule on the energy scale is an amide-ester. Glycine comes in third position, 10.8 kcal mol⁻¹ above the most stable compound. In fourth position one finds an amide-alcohol. In all cases, it is clear that the α -amino acid structure is not that of lowest energy.

This simulation of the relative stability of the first prebiotic systems is helpful in understanding why glycine has not been observed in space in the gaseous phase. In the extreme conditions, where the molecules formed are those of lower energies, glycine simply does not satisfy the criterion of maximum stability. One is led to the puzzling conclusion that the peptide bond of proteins that is thermodynamically the most stable insures a strong bonding between α -amino acids, which are not the stable structures of their families.

14.3 Conclusions and Prospective

State of the art numerical simulations have reached a level of reliability such that predictive studies can now be undertaken in a large variety of scientific fields, including exobiology. The calculation of spectral signatures is precise enough to stimulate new laboratory experiments and observational searches for small molecules in the radio millimeter range. Numerical simulations in infrared spectroscopy not only make possible the interpretation of the spectra, but also their prediction in the case of unstable species (free radicals or ions). The frequencies are obtained with a good precision but the difficult calculation of absolute

intensities needs still to be improved. Concerning chemical modeling, the numerical simulation of elementary processes proved particularly useful for testing the reaction networks and often the theoretical results led to drastic revisions of pre-conceived ideas. All this applies to the gaseous phase.

The progress to be expected in the coming years concerns chemistry at the solid/gas frontier. Solids play multiple roles. On the one hand, spectral signatures are modified by adsorption; on the other hand, it is the chemical processes themselves that are modified. Reactions that were too exothermic to form stable species in the gaseous phase (for example the synthesis of H_2) will be possible because the solid will absorb the excess energy released. New types of reactions linked to the catalytic properties of dust and ices will be considered. In exobiology, the study of the interactions between organic molecules and minerals will be a priority. This field of research will stimulate progress in numerical simulations and boost the development of new molecular dynamic methods for reactive processes taking into account the solid/gas interface.

Another aspect of primary importance is the reactivity in the liquid phase. There are several models of solvents available today. Some consider that reactants and products are embedded in a continuum with a characteristic dielectric constant; others include individual interactions with the solvent molecules, the two approaches being coupled in the more elaborate treatments. Although these methods already provide a valuable insight in the reaction mechanisms taking place in liquids, improvements are still necessary to reach the precision of gaseous phase studies.

Last but not least, the development of computing facilities will greatly determine the type of research that will be done in the near future, particularly for exobiology where numerical simulations have to consider systems of large dimensions. If the theoretical approach provides a better understanding of the elementary mechanisms that lead to prebiotic species, it should also be stressed that the emergence of the molecules of life is a formidable challenge for theoreticians. Success in the future will be in the interdisciplinary collaborations.

Acknowledgments

Writing this chapter would not have been the same without the research performed at LETMEX (Laboratory for Theoretical Study of Extreme Media) by G. Berthier, C. Dimur, Y. Ellinger, F. Pauzat and D. Talbi who participated actively in the development of quantum astrochemistry. Most of the numerical simulations have been performed on the CINES and IDRIS computational facilities. The authors would like to thank J. Reisse, H. Martin and M. Gargaud for useful comments and R. Hewins for his kind assistance with the English version.

References

General textbooks

- Daudel R., Leroy G., Peeters D., Sana M. (1983). *Quantum Chemistry*, Wiley, New York.
- Green N. (2003). *Comprehensive Chemical Kinetics: Unimolecular Kinetics*, Vol. 39, Elsevier, London.
- Parr R.G., Yang W. (1989). *Density Functional Theory of Atoms and Molecules*, Oxford University Press, New York.
- Rivail J.L. (1989). *Eléments de Chimie Quantique à l'usage des chimistes*, InterEditions/Éditions du CNRS, Paris.
- Szabo A., Ostlund N.S. (1982). *Modern Quantum Chemistry*, McGraw-Hill, New York.
- Townes C.H., Schawlow A.L. (1975). *Microwave Spectroscopy*, Dover, New York.

Specialized articles cited in text

- Abelson, P. (1959). *Science*, **124**, 935.
- Allamandola L.J., Tielens A.G., Baker J.R. (1985). *Astrophys. J.*, **290**, L25.
- Barker J., Allamandola L.J., Tielens A.G. (1987). *Astrophys. J.*, **315**, L61.
- Bartlett R.J. (1981). *Ann. Rev. Phys. Chem.*, **32**, 359.
- Becke A.D. (1988). *Phys.Rev.A*, **38**, 3098.
- Becke A.D. (1993). *Phys.Rev.A*, **98**, 5648.
- Berthier G. (1954). *C. R. Acad. Sci.*, **288**, 91.
- Bockelée D., Lis D.C., Wink J.E., Despois D., Crovisier J. et al. (2000). *Astron. Astrophys.*, **353**, 1101.
- Bogey M., Demuyck C., Destombes J.L. (1984). *Astron. Astrophys.*, **138**, L11.
- Bogey M., Demuyck C., Destombes J.L. (1986). *J. Chem. Phys.*, **84**, 10.
- Brown R. (1981). *Astrophys. J.*, **248**, L119.
- Brown R., Godfrey P., Elmes P., Rodler M., Tack L. (1985). *J. Am. Chem. Soc.*, **107**, 4112.
- Cernicharo J., Guélin M. (1996). *Astron. Astrophys.*, **309**, L27.
- Cheikh F., Pauzat F. (1999). *Astron. Astrophys.*, **348**, 17.
- Chiaromello J.M., Talbi D., Berthier G., Ellinger Y. (2005). *Int. J. Astrobiology*, **4**, 125
- Cohen M., Allamandola L., Tielens A.G., Bregman J., Simpson J.P., Witteborn F.C., Wooden D., Rank D. (1986). *Astrophys. J.*, **302**, 737.
- Cooper G.W., Onwo W.M., Cronin J.R. (1992). *Geochim. Cosmochim. Acta*, **56**, 4109.
- De Frees D.J., Loew G.H., McLean A.D. (1982). *Astrophys. J.*, **254**, 405.
- De Frees D.J., McLean A.D. (1989). *Chem. Phys. Lett.*, **158**, 540.
- DeFrees D.J., Miller M.D., Talbi D., Pauzat F., Ellinger Y. (1993). *Astrophys. J.*, **408**, 530.
- De Graaf R.M., Visscher J., Schwartz A.W. (1995) *Nature*, **378**, 474.
- D'Hendecourt L., Léger A. (1987). *Astron. Astrophys.*, **180**, L9.
- Dimur C., Pauzat F., Ellinger Y., Berthier G. (2001). *Spectrochimica Acta*, **57A** (4), 859.
- Duley W.W., Williams D.A. (1981). *MNRAS*, **196**, 269.
- Ellinger Y., Pauzat F., Lengsfeld B.H. (1999). *J. Mol. Structure THEOCHEM*, **458**, 203.

- First ISO Results (1996). *Astron. Astrophys.*, **315**, 1.
- Guélin M., Cernicharo J., Travers M.J. et al. (1997). *Astron. Astrophys.*, **317**, L1.
- Haldane, J.B. (1929) *Rationalist Annual*, p. 148.
- Hall G.G. (1951). *Proc. Roy. Soc.*, **A205**, 541.
- Herbst E., DeFrees D.J., Talbi D., Pauzat F., Koch W., McLean A.D. (1991). *J. Chem. Phys.*, **94**, 7842.
- Hinze J., Roothaan C.C.J. (1967). *Prog. Theoret. Phys. Suppl.*, **40**, 37.
- Hocking W.H., Gerry M.C.L., Winnewisser M. (1975). *Can. J. Phys.*, **53**, 1869.
- Hohenberg P., Kohn W. (1964). *Phys. Rev.*, **136**, 864.
- Hudgins D.M., Sandford S.A., Allamandola L.J. (1994). *J. Phys. Chem.*, **98**, 4243.
- Hudgins D.M., Allamandola L.J. (1995a). *J. Phys. Chem.*, **99**, 3033.
- Hudgins D.M., Allamandola L.J. (1995b). *J. Phys. Chem.*, **99**, 8978.
- Hudgins D.M., Sandford S.A. (1998). *J. Phys. Chem. A*, **102**, 329, 344, 353.
- Huron B., Malrieu J.P., Rancurel P. (1973). *J. Chem. Phys.*, **58**, 5745.
- Joblin C., Tielens A.G., Allamandola L.J., Geballe T.R. (1996). *Astrophys. J.*, **458**, 610.
- Kohn W., Sham L.J. (1965). *Phys. Rev. A*, **4**, 1133.
- Langhoff S.R. (1996). *J. Phys. Chem.*, **100**, 2819.
- Leach S. (1986). *J. Electron. Spectrosc. Relat. Phenom.*, **41**, 427.
- Lee C., Yang W., Parr R.G. (1988). *Phys. Rev. B*, **37**, 785.
- Leger A., Puget J.L. (1984). *Astron. Astrophys.*, **137**, L5.
- Lennard-Jones J.E. (1930). *Proc. Roy. Soc.*, **A129**, 598.
- Lis D.C., Keene J., Young K., Phillips T.G., Bockelée D., Crovisier J., Schilke P., Goldsmith P.F., Bergin E.A. (1997). *Icarus*, **130**, 335.
- Luine J.A., Dunn G.H. (1985). *Astrophys. J.*, **299**, L67.
- McCarthy M.C., Travers M.J., Kalmus P., Gottlieb C.A., Thaddeus P. (1996a). *Astrophys. J.*, **467**, L125.
- McCarthy M.C., Travers M.J., Kovacs A., Gottlieb C.A., Thaddeus P. (1996b). *Astron. Astrophys.*, **309**, L31.
- Miller S. (1953). *J. Am. Chem. Soc.*, **77**, 2351.
- Miller S., Urey, H.C. (1959). *Science*, **130**, 245.
- Møller C., Plesset M.S. (1934). *Phys. Rev.*, **146**, 618.
- Nesbet R.K. (1955). *Proc. Roy. Soc.*, **A230**, 312.
- Ohshima Y., Endo Y., Ogata T. (1995). *J. Chem. Phys.*, **102**, 1493.
- Ogata T., Endo Y., Ohshima Y. (1995). *J. Am. Chem. Soc.*, **117**, 3593.
- Pauzat F., Ellinger Y., McLean A.D. (1991). *Astrophys. J.*, **369**, L13.
- Pauzat F., Talbi D., Miller M.D., DeFrees D.J., Ellinger Y. (1992). *J. Phys. Chem.*, **96**, 7882.
- Pauzat F., Talbi D., Ellinger Y. (1995). *Astron. Astrophys.*, **293**, 263.
- Pauzat F., Ellinger Y. (2001). *MNRAS*, **324**, 355.
- Perdew J.P., Wang Y. (1986). *Phys. Rev. B*, **33**, 8800.
- Perdew J.P., Wang Y. (1991). *Electronic Structure of Solids 91*, Academic Press, Berlin.
- Pople J.A., Nesbet R.K. (1954). *J. Chem. Phys.*, **22**, 571.
- Roothaan C.C.J. (1951). *Rev. Mod. Phys.*, **23**, 69.
- Roothaan C.C.J. (1960). *Rev. Mod. Phys.*, **32**, 179.
- Rubin R.H., Swenson G.W., Benson R.C., Tigelaar H.L., Flygare W.H. (1971). *Astrophys. J.*, **169**, L39.
- Schutte W.A., Tielens A.G., Allamandola L.J. (1993). *Astrophys. J.*, **415**, 397.

- Snyder L.E., Buhl B., Edrich C.H. (1972). *Astrophys. J.*, **177**, 625.
- Szczepanski J., Roser D., Personnette W., Eyring M., Pellow R., Vala M. (1992). *J. Phys. Chem.*, **96**, 7876.
- Szczepanski J., Vala M., Talbi D., Parisel O., Ellinger Y. (1993). *J. Chem. Phys.*, **96**, 4494.
- Szczepanski J., Vala M. (1993). *Astrophys. J.*, **414**, 646.
- Talbi D., Hickman P., Pauzat F., Ellinger Y., Berthier G. (1989). *Astrophys. J.*, **339**, 231.
- Talbi D., Ellinger Y., Herbst E. (1996). *Astron. Astrophys.*, **314**, 688.
- Talbi D., Ellinger Y. (1998). *Chem. Phys. Lett.*, **288**, 155.
- Talbi D. (1999). *Chem. Phys. Lett.*, **312**, 291.
- Taylor P., Scarlett M. (1985). *Astrophys. J.*, **293**, L49.
- Tobita S., Leach S., Jochims H.W., Rühl E., Illenberger E., Baumgärtel H. (1994). *Can. J. Phys.*, **72**, 1060.
- Travers M.J., McCarthy M.C., Gottlieb C.A., Thaddeus P. (1996). *Astrophys. J.*, **465**, L77.
- Vala M., Szczepanski J., Pauzat F., Talbi D., Parisel O., Ellinger Y. (1994). *J. Phys. Chem.*, **98**, 8187.
- Vosko S.H., Wilk L., Nusair M. (1980). *Can. J. Phys.*, **58**, 1200.
- Weimer H.A., McFarland B.J., Li S., Weltner W. (1995). *J. Phys. Chem.*, **99**, 1824.

15 Artificial Life or Digital Dissection

Hugues Bersini

15.1 Introduction to Artificial Life

Would a biologist be surprised to be told that the computer should be able to help him or her make substantial progress in his discipline? It is doubtful. Nevertheless, proponents of artificial life would bet that it can help him read and understand the natural phenomena forming the basis of his studies through massive use of information technology, principally using simulation. By “artificial life” we primarily mean using information technology in the first instance and, to a lesser degree, robotics, to reproduce mechanisms common to living organisms and reproducing them in a non-biochemical substrate. In artificial life, as with artificial intelligence, we understate the importance of the substrate for the benefit of the function. There is absolutely no need for a particular materiality to make vital functions emerge. Life begins at the intersection of a series of processes, which need to be isolated, differentiated and duplicated as such in computers. Between the concept of the function creating the organ, which is the realm of engineers and the organ creating the function, more the sphere of chemists and physicists, artificial life principally opts for the former. While still listening to those in favor of the first, it nevertheless seeks to be engaged with today’s biology, but, if necessary, is ready to suggest a new, dissident version, one essentially centered on producing unrestrained software production and experimentation inspired by real life. Artificial life encourages its natural counterpart to free itself from its material substrate, while being characterized by a range of functional properties: metabolism and self-maintenance, individuation, environmental autonomization, self-reproduction, evolution, encoding, each of which is eager to be replaced with a computer-generated equivalent.

This movement towards “de-materialization” is so great that no supporter of this discipline would be surprised to learn that chemical material is not indispensable; chemical material is the only thing which has been regarded as capable of supporting living organisms up to now, but it is one of many possible material transpositions of real life. They would not be surprised to learn that this material could even be uselessly complicated. Although life does indeed owe this material everything, this is as a result of a simple historical contingency. It was simply there and was therefore the first to produce these functions, doing so as laboriously as it was able. Life had to push onward with carbon, because there

was nothing else available at the time. The history of philosophy, technologies and sciences can furnish many versions, which pre-figure the same process, such as Descartes and Leibnitz who considered living organisms to be nothing but simple mechanisms, the essential difference between the two being marked by conscience, and the mechanical animals of the nineteenth century, such as the Vaucanson's famous dog – the illustrious predecessor of the little Sony robot dog. This is also why Langton, the father of artificial life, sees in his creation a scientific discipline, which studies life, not as it is, but as it could have been and as it could still be, prioritizing function over substrate. We can easily understand how artificial life is linked to exobiology; this other recent scientific discipline is centered on life and the study of its origins, not only on the obvious environment of Earth, but also throughout the universe. This discipline cannot restrict itself to a mere functional view of life, in order to detect it elsewhere, as the material substrate could be something completely different, on a distant planet or a university lab. The exobiologist dreams of detecting the presence of another version of a terrestrial organism, however rudimentary, somewhere else in the universe, through its behavior, even before being able to dissect it. It is in this discipline, furthermore, that researchers are striving hardest of all to agree on the mechanisms essential for characterizing the living, as in Joyce's already outdated definition (1994): life is a self-maintaining chemical system capable of Darwinian evolution; or Clarke's more recent, more complete version (2003): all forms of life takes place in a single organism where reproduction is made possible by a collection of indispensable instructions. As for an organism, it is defined as any physical entity, which can, in an adapted environment, affect the flow of energy, which passes through it while carrying out certain functions.

For its part, artificial life does not attempt to provide a new definition of life, any more than do most biologists. It is not expected to – which scientific discipline is? The concept of *life*, as opposed to *gravity* or *electromagnetism* or the *quantum reduction of a wave packet* was already in widespread existence prior to any scientific reading or reification. It is all-purpose and has been weakened by pre-existing semantic limits, which are now out of control. This has allowed Lovelock the latitude to compare the planet Earth, which he baptized “Gaia”, to a living organism and has done the same for all the other protagonists of artificial life who tend to see life everywhere as soon as an amusing little animation appears on their screen or a mechanical dog wags its tail. The Web even becomes alive, because it is emergent, out of control, self-maintaining and never stops developing. Another difficulty, which is more than just semantic, is the extreme discomfort that science feels when it is applied to the origins of phenomena. Some of these are the origin of the universe, the origin of consciousness, and of course the origin of life. The temporal discontinuity of a phenomenon requires its precise definition, while paradoxically this same definition is often based on this discontinuity – the vicious trap of circular definitions. The rejection of a definition is often compensated for by a list of functional prop-

erties, which never finds unanimity amongst its authors. Some demand more properties, others require fewer and often different properties. Compromising does not solve these problems in any way, as each of these properties, often stated in terms of the a vague expression such as emergence, self-maintenance, metabolism, and self-replication, do not provide the same functional vision on examination. For this reason, as it grants itself no authority to give any definition, artificial life can nevertheless play a determining role by writing and implementing software programs, to remove any ambiguity from processes to which many biologists will give the same name, while seeing very different mechanisms in them.

What is new in this project? Has not biology always shared this concern about understanding living organisms? It has certainly had the same concern about understanding them, but they see no use in computer generated replication, from the crudest to the most precise if it does not allow them to understand better any of these mechanisms. The biologist does not aspire to become a sorcerer's apprentice and his eyes remain fixed on the natural world: dissection, detailed recording, attempting to understand and predict its final way of working. These models mostly remain very descriptive, linguistic, qualitative, relational, causal (these systems are made up of these components and these same components influence the others, etc.) and have not been used to lend themselves to algorithmic writing up to the present. In artificial life however, these models form the very basis of the computer programs. They are built on a basis of these objects and relations, which have up to now been known only by biologists. The computer has taken over the primary role – biology has to approach the computer, rather than the other way around. The computer has to be made to function biologically, integrating algorithmically those lessons, which seem to be essential to life and testing them by means of this computer-generated guinea pig. At this point it is important to distinguish between artificial life and what is more commonly known as bioinformatics (another recent, very fashionable hybrid of informatics and biology). However, by the latter we understand the use, again on a large-scale, of IT (Information Technology) in the handling and analysis of data gained from biological organisms, for example gene sequencing, the comparison of these sequences, the prediction of 3D protein conformation, the classification or discrimination of microarrays, etc. In this case, the IT element is simply a box full of very useful tools, very often statistical, needed for an iterated handling of a considerable amount of data and nothing else. The biological data is new, the treatment of it was unusual so far, but the computer algorithms used are very classical.

Through artificial life, the computer and biology working in tandem show what they have most in common: their ability to function at different levels of abstraction. A new dialogue is being set up between the simple and the complex, because at primary levels the mechanisms must necessarily be simple. Both can engage in an infinite number of tests, errors, and selection of what works, that as much as by this hierarchical organization in abstraction as by brute force, test-

ing multiple solutions and only retaining the best. Things that appear complex at first glance are not systematically so, and a closer reading, only made possible by the computer, may reveal an unexpected simplicity. A computer, at its ultimate level of function, the logical gates “or” and “and” become remarkably simple. The increasing complexity comes from above, by successive functional aggregation according to the engineer’s plan. Biology resembles it closely except in one fundamental aspect: only a blind engineer (to borrow Dawkins’s highly expressive image) could have assembled the pieces from below to make the higher functions emerge. By the same token, only the computer, like some new form of microscope, can allow us to discover and understand what this blind engineer designed by simple repetition at all levels of a selection process: the mechanisms on the lower story supporting the functions of the top.

Historically, the first American conference inaugurating artificial life goes back to 1987, and was closely followed by European versions of these same conferences, known as ECAL (European Conferences on Artificial Life), the first one taking place in Paris in 1991 and the second in Brussels in 1993. Though it portrays itself as innovative, this discipline can be seen to have multiple origins. The first was the inability of artificial intelligence to come to grips with the problem of how a sensory-motor interface recognizes the world (how a robot sees and acts in its world), as long as it still kept its distance from biology – the materialization of the cognitive apparatus. However, the robot in us cannot be explained by neglecting the essential parallel and adaptive elements of our brain. By overplaying the function card at the expense of the substrate, artificial intelligence, doubtless the victim of its own success has gone a bridge too far, to the point of completely denying the importance of the complexity of the neural cabling, its extreme parallelism and its plasticity. By taking a big step backwards, artificial life is getting used to reincarnating the cognitive functions, both in the real biological world and in the environment in which they are used and express themselves. It chooses to inscribe “robotic cognition” in “robotic life”, so that robotic cognition is only created as a support for and product of a robot exploring and using its environment in order to remain viable in it. Man is a reed, which bows and bends in the world surrounding it, before becoming a thinking being and who, by virtue of his cognitive projections, is able to fight this environmental pressure or use it for his own ends.

A second origin is this interest for non-linear and self-organizational phenomena, which can be seen in physics and other branches of science. It fundamentally calls into question the procedure of dividing a mechanism into its parts in order to understand how these parts collectively work. While this can only be understood as a close meeting of parts the way in which they are used, taken separately is only one aspect of the way they collectively work, as this reciprocal influence can only be understood by putting the parts in contact with each other via a computer simulation. The non-linear phenomena cannot be divided up or added together: they interpenetrate into each other, making it impossible to understand their combined behavior simply by looking at either of them

separately. At this point, the classical mathematician can contribute nothing and is unable to provide any remedy that helps to understand how these new properties emerge when the different parts meet and work in tandem. These mathematicians often limit themselves, titillated by some phenomenon revealed by IT, setting off like Offenbach's cavaliers, to extrapolate the number of parts ad infinitum so as to obtain a reading of the *average* behavior of these systems by statistics. Although this understanding has the advantage of being formal, it is still a caricature. Where a formal approximation is possible, it only says something in very special cases, restricting itself to reproducing the IT element. It cautions and solidifies, but reveals nothing really new. Finally we should also be able to see something of a revival of good old cybernetics and systemics of the 1940s and 1950s in artificial life (see Langton 1989 for a short history). It has however been given a completely new impetus by the availability of cheap and powerful computers. Often, what is found in this so-called new discipline is simply re-formulated information from that period, rehashed and updated by the use of IT or robotics. The language appears more complete, but in fact it is simply more algorithmic and verifiable.

The bulk of the work carried out in artificial life is based on lessons learned from life. I shall define them as functional emergence, self-maintenance, closure and individuation, growth, self-replication, evolution, plasticity and adaptability, environmental autonomization through significant integration, genotypical coding of the phenotype, and, despite their rather sibylline nature, they will be the titles of the following chapters. Before this however, beginning at the next chapter, I shall attempt to set out the history of life as the disciples of artificial life understand it, by placing these different lessons on a temporal and causal axis, showing which one is indispensable to the appearance of the next. The new task of the IT specialist is to set up experimental software platforms (such as networks, cellular automata or genetic algorithms) where these different lessons, whether taken in isolation or together, are tested, simulated, and, more systematically analyzed. The project will be all the more vigorous as these mechanisms are relevant, at a given level of abstraction, in a multitude of biological systems: genetic, neural, hormonal, immune, cellular, animal, etc. More effective experimentation and understanding of each one of these could have a wide impact if it is extended to all the biological disciplines out of which these mechanisms are abstracts. Furthermore, as we will show in detail, this functional reading of the biological processes also underpins a new way of systems engineering, which owes much to parallelism, emergence and adaptability.

Whatever may be understood from the preceding paragraphs, the biologist obviously remains the most important partner of the discussion; but what may be expected from this artificial life? What can the biologist expect from these new "Merlin hackers", whose ambitions seem above all disproportionately naive. Christian de Duve (2002), holder of the Nobel Prize for Medicine and famous more recently for his work on the origin of life, recently stated:

“On their computer platforms, theoretical biologists replace molecular structures with symbols and chemical reactions with algorithms. Practices of this kind may prove useful and lucrative. Nevertheless, calling this artificial life is nonsense. If life is ever recreated artificially, this will be in a test tube and not in a computer.”

I would answer that computer platforms, irrespective of whether they are alive, could be useful in several ways, presented in terms of their increasing importance or by force of impact in the rest of the introduction. First of all, they can open the door to a new style of teaching concerning the major biological ideas. This was the case, for example, for Richard Dawkins (1989) who, bearing the Darwinian good news, did so with the help of a computer simulation where creatures known as “biomorphs” evolve on a computer screen by means of genetic algorithms. There is nothing new here, apart from an unsurpassable illustration of Darwinian principles. The fact that ever more surprising and complex creatures appear in a deliberately simplified succession of selection, reproduction and mutation does not explain anything new, but shows that this process works and works well, acting as a new operational proof. It also acts as a way of opposing a whole vigorous anti-Darwinian intellectual community that contests the role of chance and selection as the sole mechanism producing the extreme complexity of living organisms. Their arguments are often guilty of major errors of logic that neglect the parallelism of genetic tests, the time available, and the constructional and progressive aspects of the genetic mechanisms employed. This is also the purpose of many ecosystem or vegetable growth software simulation systems, which teach the biologists nothing, except once again: it works like this, or it could work like this and furthermore, we’ll show it to you. If a picture is worth a thousand words, this is all the more true of a computer simulation, especially when it is highly colored and appears on a screen. Certain biological phenomena, such as the appearance of complex or functional structures by evolutionary iterations from a basis of simplistic structures, which have no initial significance, are illustrated in this way. Only informatics can reproduce a near infinity of elementary mechanisms in a confined space and time and reach the result in the required time.

These platforms and simulations can, insofar as they are sufficiently flexible, quantifiable and universal, be used precisely by the biologist, who will find in them a simplified means of simulating and validating a given biological system. They have similarities with other computer-generated environments, highly valued amongst engineers or other scientists, who use them as an aid to their experiments and models. I am thinking in particular about a MATLAB or Mathematica for biologists. Cellular automata, Boolean networks, genetic algorithms and algorithmic chemistry are excellent examples of software to parameterize and use to produce the natural phenomenon required. Their predictive power varies from very qualitative (their results show very general trends in the real world) to very quantitative (the numbers produced by the computer may be compared to those in the real world that we are seeking to model). Although they are very

qualitative, due to the fact that they need to be translated into an algorithmic structure these programs often allow us to examine those mechanisms known to be responsible for observed patterns of behavior, as this structure is already the guarantee of an advanced understanding accepted by all. Algorithmic writing is an essential stage in formalizing the elements of the model and making them objective. The more the model allows us to integrate what we know about the reality reproduced, the detailed structures of objects and relationships between them, the more the predictions will move from tendentious to precise and the easier it will be to validate the model, according to Popper's method, the way in which physicists wish to see all evolving biology models.

Although algorithmic writing is less demanding than mathematical writing, it requires a great degree of rigor and thus a much sharper clarification of various scientific concepts than is found in biological literature in versions as quite ambiguous. This algorithmic analysis cannot work without effort and requires biologists to explain any background information such as emergence for example, or genetic coding, which they may view as obvious even if it is extremely subtle. New original mechanisms may be discovered, as it is their multiple iterations in time and space, only made possible through the computer, which allows us to understand how they underlie the observed behavior. Ideally, through systematic experiments, these platforms can lead to the discovery of new natural laws, whose impact will be greater than the simulated abstractions present in many biological realms. Thus in the 1950s, when Alan Turing discovered that a simple diffusion phenomenon, propagating itself at a different speed, depending on whether it was subject to a negative or positive influence, produces zebra or alternating motifs, it had a considerable effect on a whole section of biology studying the genesis of forms (animal skins, shells of sea creatures). When some scientists discover that the number of attractors in a Boolean network or a neural network has a linear dependency on the number of units in these networks (Kauffman 1995), these results can equally apply to the number of cells expressed as dynamic attractors in a genetic network or the quantity of information capable of being memorized in a neural network. Entire chapters of biology (neural, genetic, protein, immunohormonal) had to be re-read or even re-written in the light of these discoveries. Biology is affected when scientists observe a non-uniform connectivity in many networks, whether social, technological or biological, showing a small number of nodes with a large number of connections and a greater number of nodes with far fewer, and when, in addition, they explain the way in which these networks were constituted in time (Barabasi 2002).

As several scientific philosophers have noted, the strength of a scientific model essentially results from its power to allow inferences to be drawn, from its capacity, almost independent, to produce new knowledge. The word *almost* is fundamental here, because although this model has the force of law in its determinist version and demands to be applied scrupulously, a scientist is still indispensable to its application. Although fact B certainly follows from fact A, still a scientist is needed to execute that inference. Until the advent of IT to science,

man, objective as he might have been, but despite that, was required for the manipulation and execution of the model. Ever since scientists have been able to program and transpose their models to form a computer program, the extraordinary consequence has been to transform this inferential mechanism from *almost independent to totally independent*. As the computer has integrated this inferential power into its circuits, it can, without any human holding its hand, start predicting the future. The computer model substantially thinks for itself and it can not only carry out new tasks and produce new things in a completely de-humanized fashion, but it can expand the language that can be used for producing inferences. Computer language, although very rigorous, offers more flexibility than any mathematical language. Finally, the icing on the cake is that it is able to produce inferences which, in practical terms, far surpass those possible by the human mind due to its calculating power. It is incredibly fast, needing one nanosecond to carry out an addition and is capable of taking far more factors into consideration when producing a model than the human mind could ever consciously process.

Provided that one knows a programming language, algorithmic structure is more explicit, simpler to analyze and more adaptable than mathematical structure. Given this speed of processing, the computer can replay certain scenarios of biological evolution that have taken place over millions of years endlessly, without the programmer having to resort to gnawing at the mouse. This allows the scientist to test several hypotheses, retaining only that one which resembles the current situation most closely. The programmer creates new worlds, which evolve on their own and can, as necessary, select those projects, which are worth allowing to evolve. The computer suggests a result and the scientist adapts to it, looking to understand the result and ensuring that it is not a simple artifact linked to the intrinsic limitations of the method of inputting and processing digital information. Seen in this light, IT and, to an even greater extent, artificial life, are of great benefit to biology.

We are therefore thinking about a new algorithmic language, as useful to biology as mathematics is to physics; a language that could easily express biological realities, allowing us greater understanding, and whose workings, when exposed, would already carry a whole range of universally applicable inferences and predictions. Artificial life is of course at its apogee when it allows us to reveal new biological facts, to destabilize biologists' pre-suppositions or generating new knowledge, rather than simply illustrating or refining the old. At the moment, the fact that this discipline is new and has a certain immaturity in comparison with biology, its illustrious elder and predecessor, has lead several observers to remain in want of information in the face of the current discrepancy between promises and reality. In my opinion they have tended to underestimate the importance of the results already obtained, as they are too attached to their microscope. They have lost the perception of the forest by being drowned in the study of the trees. In order to appreciate the results, they must show less reticence and arrogance towards these new computer explorers; they must show

more curiosity and conviviality for these people who, like them, have set out on the conquest of life, but by using methods, which attract less attention because they are electronic only.

The last of our natural partners, who is of slightly less interest to us here, is the engineer. Over the whole of time, nature has built up a reserve of inexhaustible inspirations, which have helped make useful artifacts for man. Artificial life has led to new computer tools, such as genetic algorithms, Boolean networks, robots, which learn by experience, cellular machines and others, which create a new vision of IT for the engineer: parallel, adaptable and autonomous. In this kind of informatics, complex problems are tackled with the aid of simple mechanisms, but infinitely iterated in time and space. In this kind of informatics, the supreme paradox is that the engineer must be resigned to partly losing control if he wishes to obtain something useful. The computer finds the solutions by trial and error, while the engineer concentrates on observing and indicating the most promising directions for research. He simply tells the program whether it is *hot* or *cold*. He is still indispensable because he can appreciate the progress that the computer has made towards the solution without having the slightest idea how it achieves it. He points out some promising directions but is incapable of running over the billions of possible pathways to locate the one which will really lead to the goal. This is the case for the intensive use that engineers have made of artificial neural nets or genetic algorithms. The computer or robot game industry has in the main found its sources of inspiration in artificial life. Hollywood film producers are also highly indebted to it, using algorithms of collective animal behavior to create either virtual images of moving flocks of real or imaginary animals, such as the dinosaurs in Jurassic Park or the passage of shoals of fish or swarms of insects in cartoons. The behavior of each member of the flock is not manually considered or programmed, one by one, each in turn, but produced automatically, as the result of an algorithm for collective behavior, inspired in every way by real animals, where the behavior of each one depends simply and solely on its neighbors.

The following chapter will attempt to sketch out a history of life in successive strata; this history will certainly be very incomplete and full of unknowns, but most people involved in artificial life will be in agreement. Then, many of the lessons set out in this functional history will be reconsidered and examined in depth in the following chapters on functional emergence, self-maintained chemical network, self-replication, plasticity and adaptability genotypical and phenotypical coding. The chapter aims to provide a fairly complete panorama of events in artificial life and points out the personal contributions and work carried out by the IRIDIA laboratory. A proviso is necessary here; choosing to give a wide coverage of the different types of IT and robotic developments in artificial life involves a large number of descriptive shortcuts, the idea being to make the reader aware of this work rather than to provide him with a complete description. It is then up to him to widen his understanding of one or more of these developments by additional reading.

15.2 The History of Life Seen by Artificial Life

15.2.1 Appearance of a Chemical Reaction Looped Network

In order for a system to emerge and maintain itself inside a soup of molecules that are potentially reactive and contain very varied constituents (which could correspond to the initial conditions required for life to appear), this reactive system must form an internally looped network, as is shown in Fig. 15.1. Above all, in order for life to begin, all of the constituent components must have been able to stabilize themselves in time. These networks of chemical reactions are thus perfect examples of systems, which, although heterogeneous, are capable of maintaining themselves indefinitely, despite the shocks and impacts, which attempt to destabilize them. This comes about through a subtle self-regeneration mechanism, where the molecules end up producing those molecules which have produced them. This is a simple system of loop production, with the producers becoming the products and the products becoming the producers: a closed system. This may be obtained on a simple level in a perfectly reversible chemical reaction, but it can be obtained more subtly, as is shown in the figure, if we let in a series of intermediary molecules before admitting those which close the loop. By this reactional roundabout in which they all participate, all molecules contribute to maintaining themselves at a constant concentration, compensating and re-establishing any disruption in concentration undergone by any one of them. If one of the molecules produced by the network does not participate in the production of any of the elements in the network, the whole network collapses as if the molecule has devoured it. The bigger the network, the more stable it is and the more molecules it will contribute to maintaining it in a concentration zone, which will vary very little despite external disruptions.

A network of this kind will be materially closed but energetically open if none of the molecules appear in or disappear from the network as a result of external factors. Energy, on the other hand, originating in external sources, is necessary for the reactions to start and take place. As soon as one of the molecules is produced in the network without, in its turn, producing one of the molecules making up the network, it absorbs and thus destroys the network. In the presence of molecules of this kind, the only way of maintaining the network therefore is to feed it materially (also indicated in Fig. 15.1). The network acts on the flow of material as an intermediate stabilization zone, made up of molecules, which may be useful to other vital functions (such as the composition of enclosing membranes or self-replication), as we will see in the following chapters. It transforms, as much as it keeps alive all the chemical agents, which it recruits. Without going into too much detail, we can say that biologists generally agree that a reactive network must exist prior to the appearance of life; it is open to external influences in terms of matter and energy, but necessarily contains a series of loops, which are most often designated as metabolism or

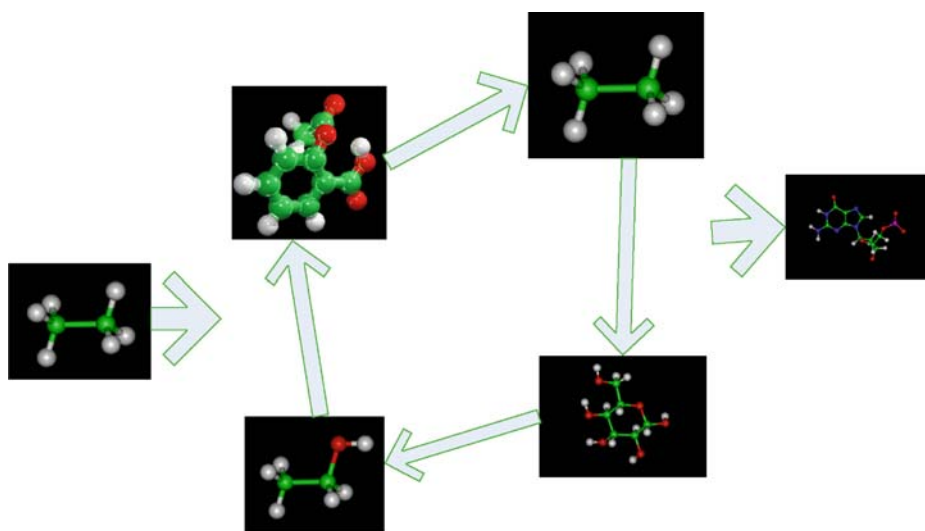


Fig. 15.1. A chemical reaction looped network, perturbed by an inlet and outlet flow of chemical material

protometabolism (De Duve 2002; Morange 2003; Ganti 2003; Maynard-Smith and Szathmari 1999).

There are a number of reactive cycles well known to biochemists, such as reactions involved in cellular respiration, during which the extremely high energy reaction of burning carbon by oxygen allows the creation of ATP molecules. These very unstable molecules lead the energy to where it is lacking, transforming themselves into much more stable ADP molecules by liberating the excess energy. The ADP molecules transform themselves back into ATP during the carbon combustion in order to close the reactive loop. During the high energy change from ATP into ADP, the energy released is used to help reactions, which are difficult to start, but are essential to life, such as protein synthesis and the duplication of DNA. Both the Krebs cycle and the reactive network involved in the glycolysis are great classics of biochemistry. We will discuss the work on these reactive networks carried out at IRIDIA in Sect. 15.3.1.3, but we can already say that it is interesting to note that these networks are balanced systems, constituting attractors in the physical sense and quite easy to obtain. Both the very nature of the chemical elements, which constitute them and the concentration of the networks seek to find equilibrium. The importance attributed to these closed reactive networks was anticipated in several works of systemic nature. For example the definition of life proposed by Perret in 1952 (see Morange 2003): “Life is an open system, capable of self-maintenance, made up of organic chemical reactions, catalysed step by step at an almost ambient temperature by complex and specific organic catalysers, which are themselves produced by the system.”

15.2.2 Production by this Network of a Membrane Promoting Individuation and Catalyzing Constitutive Reactions

The appearance of a network of this kind undeniably creates the stability necessary for utilizing its constituents in many reactive systems, together with the basic ones dedicated to the construction of membranes or the replication of molecules carrying the genetic code. This network also acts as a primary filter as it can accept new molecules within it, but can equally well reject other molecules seeking to be incorporated within it. They will be rejected, as they do not participate in any of the reactions making up the network. Can we see a primary form of individuation in this network? No, because by definition it can only be unique as, a priori, nothing allows it to be distinguished from another network. Although it is roughly possible to conceive of an interpenetration of several chemical networks, establishing a clear separation between these networks would remain a problem.

It would seem fundamental that one living organism of any kind can be differentiated from another. We know that the reproduction of a second organism from a first is a central mechanism characteristic of life and can only operate if the clone produces something to distinguish itself from its original. The best way of completing this individuation successfully and to be able to distinguish between these networks is to revert to a spatial divide, which can only be produced by a form of membrane capable of circumscribing these networks in a given space (as illustrated in Fig. 15.2). Biochemists are well acquainted with an ideal type of molecule, a raw material for these membranes in the form of lipids or fatty acids, the two extremities of which behave in an antagonistic fashion: the first hydrophilic, attracted to water, and the second hydrophobic, pushing it away. Quite naturally these molecules tend to assemble in a double layer (placing the two opposing extremities opposite each other), a double layer formed by the molecules lining up and finally adopting the shape of a sphere to protect the hydrophobic extremities from water. Like soap bubbles, these lipid spheres are semi-permeable and imprison the many chemical components trapped during its formation.

In comparing living organisms with autopoietic systems, Varela et al. (1974) insist that this membrane should be produced by the elements and the reactions making up the network (for example lipids would come from the reactions of the network themselves) and this in return promotes the emergence and self-maintenance of the network. The network can help with the appearance of the reactive and growing network by the frontiers that it sets up, the concentration of certain molecules or by acting as a catalyst to some of the reactions by its geometry or its make up. Basically, autopoiesis requires a cogeneration of the membrane and of the reactive network, which it walls up. The network presents a double closure: one chemical, linked to the looping chain of its reactions and another physical, due to the frontiers produced by the membrane. The whole, interactive metabolism and membrane pre-figures a minimal elementary cell, which already seems capable both of maintaining itself and detaching itself from

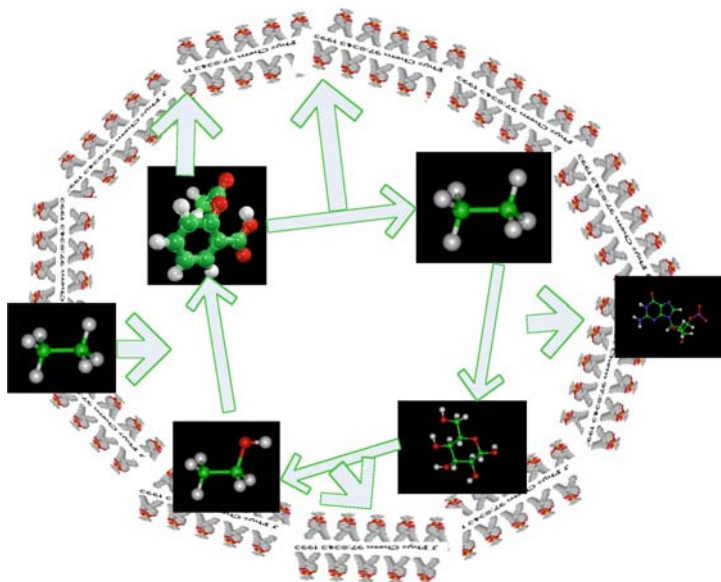


Fig. 15.2. The same chemical reaction network as in Fig. 15.1, now circumscribed by a lipid membrane and forming as such a protocell

its environment and from cells similar to it. It is at this stage of our successive conceptual additions on the way to establishing a better and more exact characterization of life that the definition given by Luisi (2002) takes on its full meaning (restating the idea of autopoiesis in more biological terms): “Life is a system that can be self-maintaining by using external energy and nutritional sources to the production of its internal constituents. This system is spatially circumscribed by a semi-permeable membrane of its composition.”

15.2.3 Self-Replication of the Elementary Cell

Self-replication, or the ability of a system to produce a copy of itself on its own, is one of the essential characteristics, which has most intrigued and impassioned followers of artificial life, beginning with the second founder of computing (von Neumann, after Turing, as we shall see later). For many biologists, if we ignore the intermediate steps described in the preceding chapters, the appearance of the first replicator constitutes the fundamental bifurcation, giving rise to the first living organism. For them, life is not simply non-dissociable from but also essentially reducible to this capacity for self-replication, as opposed to less publicized biologists such as Varela, Luisi, Ganti, Maynard-Smith and others who insist that this capacity for self-replication is not primary and for the most radical of them, not even an essential characteristic of life. In any case, it is for those supporters of the case of the self-replicating molecule being a manifestation of

the first living entity, to explain how it can actually reproduce itself without entire pre-existent cellular machinery.

On the basis of the elementary cell introduced in the preceding chapter, and in the interests of an unbroken narrative, let us imagine a short scenario leading to self-replication. The closed circuit of chemical reactions could be destabilized by some kind of disturbance, causing a growth in concentration of some of its constituents, including those involved in the formation of membranes. The membrane and the elements, which it captures begin to grow (as illustrated in Fig. 15.3) until they reach the fatal point where the balance is upset. This is followed by the production of a new cell produced by and from the old one. When the new one comes, it quickly grows fast enough to catch up with the generator and nursing cell, as a chemical network is capable of some degree of self-regeneration due to its intrinsic stability; each molecule looks around for another that it can couple to. This reconstitutes the natural chain reaction of the whole. The new membrane and the new chemical network reconstitute on their own by helping each other.

So even the most elementary diatomic molecule ($A \leftrightarrow B$) is already capable of a form of self-replication, as the following diagram illustrates. We see

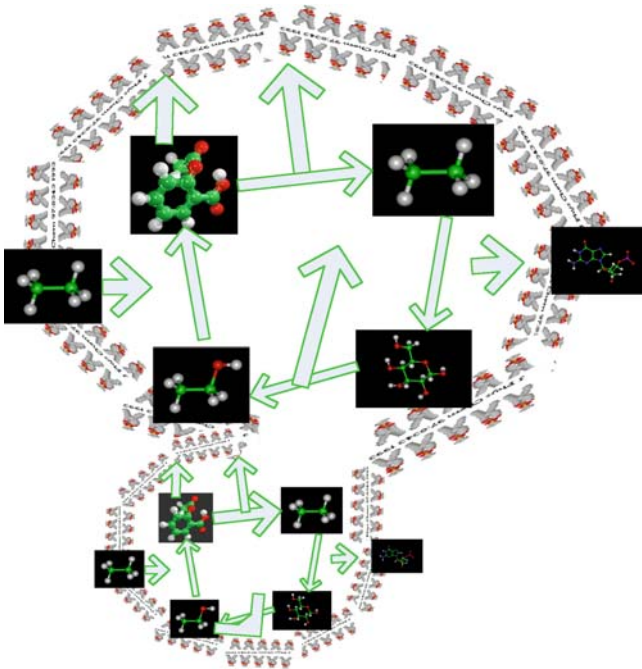


Fig. 15.3. The elementary protocell in a process of self-replication induced by the growing and the division of the chemical metabolic network together with the membrane individuating the cell

that A seeks to attach itself to B and B to A. However, rather than this elementary form of chemical self-replication coupled to physical self-replication induced by the growth and division of the membrane, life has opted for a more sophisticated physicochemical form of self-replication, more promising for the evolution to come: self-replication by the interposing of an information matrix. Each element of the matrix can only couple itself with one complementary element, these elements as a whole naturally causing the replication of the complete chain when they form part of a chemical chain. Thus these two chains have only to face each other for them to replicate themselves in concert at one go and by an identical mechanism. In the same diagram (Fig. 15.4), which is already simplified as it is a continuation of the previous one, we see an illustration of the biological alternative chosen for living organisms to self-replicate. The $A \leftrightarrow B$ part now constitutes the information matrix from which its complement, $A' \leftrightarrow B'$ can form itself and reciprocally.

In biology, it is the extraordinarily emblematic double helix of DNA that acts as a matrix, shouldering the major role in the history of life, that of the first known replicator. Our elementary cell, as shown in the diagram, must now equip itself with this information matrix, which we reduce in the figures to a little spring. Since the 1950s, Ganti (2003; a largely misunderstood researcher, but nevertheless a model advocate of and important precursor of artificial life) a Hungarian biologist, has been suggesting a first minimum system, named chemoton (see Fig. 15.5). This is the first abstract protocell, original ancestor of living organisms and possessing the three chemically-linked subsystems: a metabolic network, a membrane and an information matrix responsible for self-replication. They all depend on each other for their existence and their stability. The whole thing is theoretically capable of a whole self-replication but many important questions about the three subsystems remain to be asked, such as how the self-

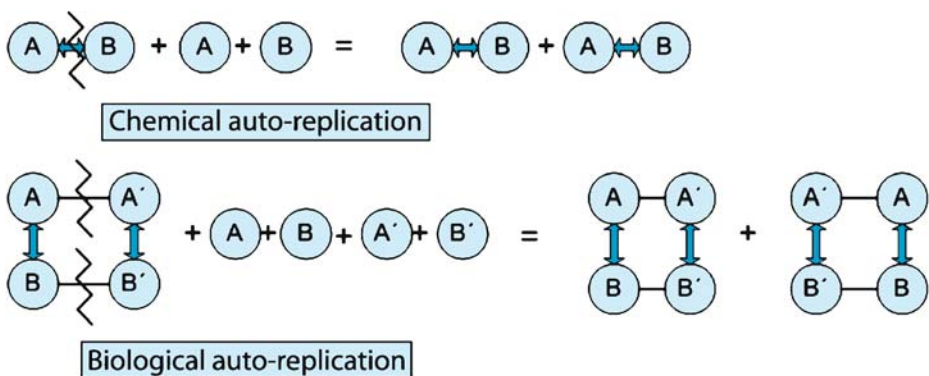


Fig. 15.4. Illustration of the biological self-replication and its differentiation from a possible chemical counterpart

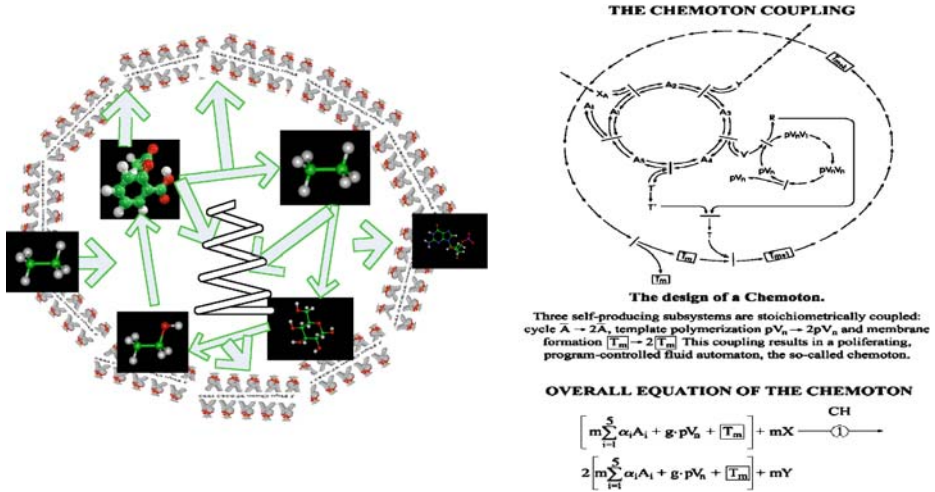


Fig. 15.5. Update of our protocell now equipped with its self-replicating informational matrix and, on the left, the schematic representation of Ganti’s chemoton (Ganti 2003)

replication of the matrix automatically accompanies the self-replication of the whole.

15.2.4 Genetic Coding and Evolution by Mutation, Recombination and Selection

These questions are so much more delicate as the information matrix introduced in the last chapter can in no way be described as information by chance, given, that apart from its self-replication, each letter constituting it contributes to the code of a functional component essential to the cell and designed on the basis of that code – the protein (as illustrated in Fig. 15.6).

Each one of these proteins, which are extraordinary cellular tools, is called to play a key role in the genesis, maintenance and functioning of this cell. One of the protein’s major and rather paradoxical roles consists in assisting the replication and decoding of the genes, which give birth to these proteins (a small-scale but often debated and controversial biological paradox of whether the chicken or the egg came first). One of the first codes very familiar to all informatics specialists is ASCII. This associates each letter of the alphabet with a binary code, the only language that the computer understands to this day. For the IT specialist, a code has two different meanings. The first, such as for the ASCII code, or the translation from base ten to base two, is a mechanism for translating a language understood by a first partner (in this case the computer) to another, understood by the second partner (in this case the human user). The code forms a bridge between two semantic worlds. This has the same meaning as when biologists describe the genetic code associating an amino acid with each

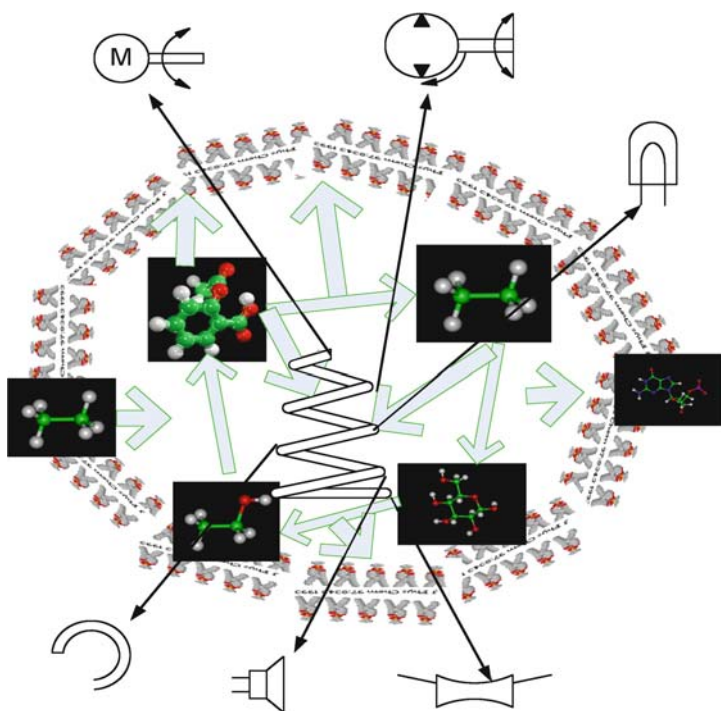


Fig. 15.6. Genetic coding of cellular tools

basic triplet making up DNA, the whole of which makes the protein. We owe the second meaning to Turing, because since that time every program executed within a computer does so by respecting and following the elementary instructions coded in the program step by step (a computer's sole reason for existence being to carry out as many programs as possible). The computer is a universal machine, as it is capable of carrying out an infinite number of programs. A complete mechanical environment is necessary as every coded instruction involves bringing certain parts of the computer into operation. Each of these instructions is in fact simply a procedure for indexing and selecting the electronic circuitry fulfilling the coded function. A specific environment is also needed for a protein to carry out that function, a function that is dictated by its amino acid composition. Both meanings of code appear to be equally applicable in the case of the genetic coding.

Biology, and in particular this faculty of self-replication, intrigued von Neumann to the extent that he devoted the essential part of his final research to it. For if we want to compare a cell to a computer and a genome to a code form, we need to explain how the computer itself was able to be created from this code. Let us follow the reasoning of von Neumann step by step, as it is the perfect illustration of a pure artificial life type of reasoning. In fact, to a sequence of

purely functional questions and showing an almost complete ignorance of biological reality, it finds a logical solution, the content of which retraces astonishingly closely those lessons we have since learned about the way biology functions. Von Neumann works on the principle that a universal constructor C must exist, who, based on the plan of some kind of machine P_M (P the plan, M the machine), must be capable of constructing the machine M^P . This idea may be simply translated by $C(P_M) = M^P$. The question of self-replication is then raised: is this universal constructor capable of constructing itself? In order to do so, the constructor must, following the example of other construction products, have a plan of what it wants to construct; in this case, correctly speaking it is the constructor's plan P_C . The problem is then expressed as follows: can $C(P_C)$ give $C(P_C)$ in order for it to be a perfect replication of the original? Von Neumann therefore realizes that the question at issue is that of the result of the construction plan, because if the constructor constructs itself, it has to add the plan itself to the product of the construction. The constructor proposes allotting two tasks to the universal constructor; first, that of constructing the machine according to the given plan and then adding the original plan to this construction. The constructor's new formula then becomes: $C(P_M) = M^P(P_M)$. If the constructor applied itself to its own plan, this time the replication will be perfect: $C^P(P_C) = C^P(P_C)$.

The fascinating aspect of von Neumann's solution is that he anticipates the two essential functions that, as we have since discovered, are the main attributions of the protein tools constituting the cell: constructing and maintaining this cell, but also duplicating the code in order for this construction to be able to prolong itself for further generations. Starting with DNA, the whole protein machinery first of all builds the cell then, by an additional procedure, duplicates this same DNA. Von Neumann did not stop at duplication, because, at the same time he imagined how this same machinery could evolve and become more complex as a result of random mutations taking place while the plan recopies itself. Since Darwin and afterwards through all evolutionary science, we have had a good idea of what the last chapter of the history of life would be, and doubtless what has stimulated most of the developments in artificial life (primarily from the point of view of engineering). This is due to the fact that the genetic code can evolve through mutation and sexual crossing between the machinery, evolving so as to produce machines that are more and more efficient.

Over the past twenty years, many of those researching artificial life have taken the idea of a genetic code to be an accepted fact, a solid and exceptionally productive base. They have been eager to show how beneficial this idea is for researching and discovering sophisticated solutions to complex problems. This research can take place though a succession of mutations and recombinations operating at the level of the code, with the best solutions proposed being preserved in order to be used for a new cycle of these same operations. A quasi optimal solution to the problem is at the end of these successive iterations, which combine mutations, recombinations and selections. The brute force of the computer is used to its full effect. They have also shown, as has Dawkins (1989), one of

the most aggressive and extreme advocates of the Darwinism, how powerful this mechanism is and how useful it would be to separate genotype and phenotype in coding and producing solutions, as does nature.

15.3 Functional Emergence

One of the usual characteristics of a complex system is that it functions at several levels of abstraction. Each level is self-sufficient, has its own levels of control and its own rules of behavior, provided that its ability to function only depends on the level immediately below. When a child plays football on the computer and selects the players of the national team, there is no doubt that the selection of the player is the result of a highly complex electronic process, where current may or may not circulate in a frightening maze of electronic gates. However, the way in which each of these logic gates works individually, remains childishly simple. We wish to declare that the selection of Zidane in some ways emerges from this intense electronic traffic. This is both true and false at the same time. It is true that the rules of football are practiced independently of the levels immediately below. It is also true that the final aim of one type of behavior, sufficiently interesting to keep the child's attention for days, is ultimately dependent on interaction between electronic gates, an otherwise more limited interest. However there is nothing emergent about this, as from this highest level to the lowest electronic level, everything is under the control of the engineer. Moving from one level to the level just preceding it is entirely programmed, planned and predictable. A very small number of engineers work on the wiring of this kind of electronic circuitry today. Others, among the best, interact with the first level of abstraction, made up of the elementary instructions of the processor. Most of them, among those who practice artificial life, are interested in the next level: programming the computer using the programming languages, programs that an automated translation system will re-write for the level just below. Even more are working at a still higher level of functional abstraction, such as when we command or set the parameters for making of applications such as with Microsoft Word or Excel work, or statistical or biological tools, such as programs for comparing genetic sequences (which are well known to biologists). This principle of functional abstraction is the key to the working of computers; it provides its richness and gives this machine this universality that Turing claimed for it.

As is shown in Fig. 15.7, computers and the different biological systems have these different levels of functional abstraction in common. They can both display new and sometimes complex functions at a higher level, although the level just below this is carried out through multiple interactions and iterations of more simple processes. At what point then, can we talk of functional emergence? This expression will be reserved for the time when the researcher discovers an original system perfected by nature in order to imitate the computer, i.e., to produce new global properties and functions at a higher level of observation from subjacent mechanisms, which have no other direct involvement at this level. As

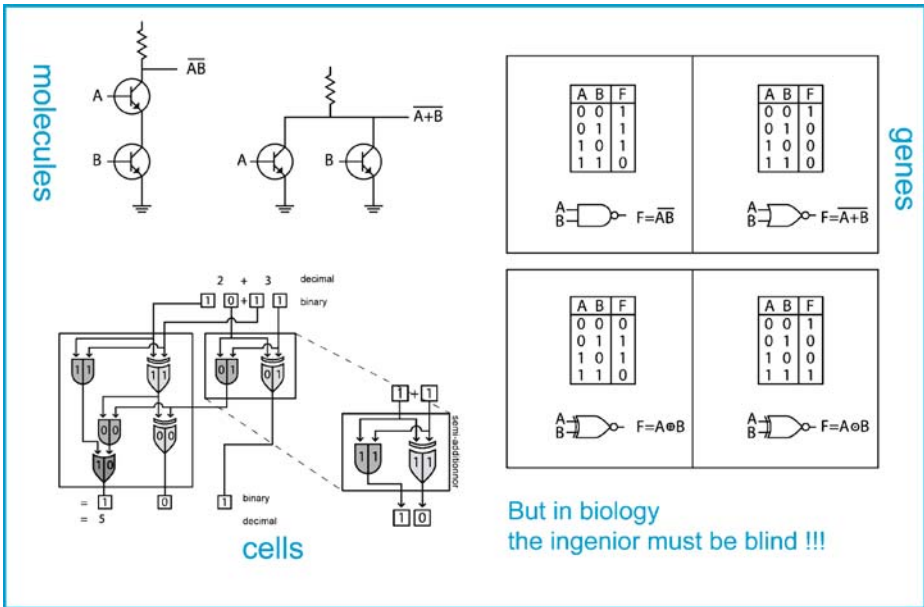


Fig. 15.7. Three levels of abstraction of a computer working. At the *top left* a transistor, which when they are grouped together (*top right*) give logical gates. Logical gates, in turn, which when grouped together can execute a binary adder. For the sake of argument, a biological equivalent is associated to each of these levels

opposed to a perfectly governed dependence between the different levels with the engineer going top-down in the quest for emergence where it is an almost archaeological method, this time bottom-up where we consider how to clarify and decode nature’s original mechanisms, so that behavioral rules, which have been perfectly dissected at one working level, give rise to another at a second level. The computer is the ideal companion for this archaeological process; it plays an essential role, like that of a microscope, allowing the rules of the system at a lower level to be revealed, so as to explain behavior in the higher spheres.

Figure 15.8 shows how a colony of ants, which has to choose between two paths to link two sources of food will always end up choosing the shortest route. There is no guiding ant and none of them is capable of identifying the shorter of the two routes. The selection of this route is an emerging property arising from many ants repeatedly executing two elementary rules; depositing pheromone on the path at a constant rate and, favoring the path that has the most of this substance. At IRIDIA, Dorigo used this idea behind the emergent phenomenon to complete one of the most successful algorithms for obtaining the shortest path in graphs, e.g., the problem of Internet routing (Dorigo and Di Caro 1999). In this problem Internet packets decide which direction they have to follow at each node in the network, based on the intensity of the track left by the packets,

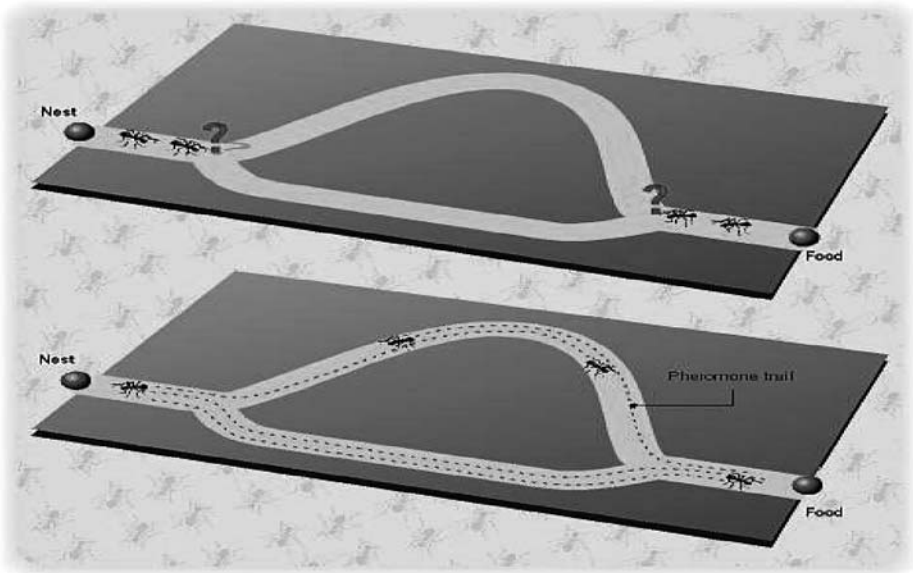


Fig. 15.8. How a colony of ants finds the shortest path to connect the food source to the nest. It is easy to understand why it is the shorter of both paths, which rapidly concentrates most of the pheromones. This phenomenon naturally self-amplifies by positive feedback since the more the path has *odor* the more it is taken by ants, and the more ants take it the more it has odor

which have circulated thus far. Using the ants as an example, the data packets leave a kind of substance behind them, telling the other packets the length of the paths traveled, the intensity of the substance being inversely proportional to the length of the path. Little by little, choosing between the two paths, the packets select the one which brings them to their goal quicker, and end up concentrating on the shortest route. This algorithm also has the immense advantage of being completely adaptable, which is an important element in a network with an essentially non-stationary load.

Followers of artificial life have realized that IT and all these macrosystems, e.g., people, towns, traffic, forests, the animal kingdom and so on, conceal a hierarchical succession of levels of abstraction and that the computer is becoming the only, not to say ideal, means of exploring these macrosystems. The ability to cause a road network to be jammed, a plant to live, a brain to think, a single bee to fly, a swarm of bees to fly, to build an anthill, to bring down an entire stand in a football stadium or a suspension bridge, is clearly the result of a collective action of those elements making up each of these systems. Referring to the previous chapters, which show the essential stages of the history of life when it is written by artificial life, this also applies to the capacity of chemical networks to stabilize itself through an internal loop, to a membrane

closing in on itself in order to imprison the other members of the network, and to their collective ability to replicate themselves. At the first level of analysis, each element possesses its own regulatory mechanisms, its own behavioral rules, but for those who are able to observe it from higher up, and gain a total perspective from a superior level, a novel and surprising spectacle is to be seen. These newly-observed faculties or functions emerge from the first level. Anyone looking through a magnifying glass will see this new behavior emerging. Where emerging is the right word, because it is undetectable amongst the elements, which make up those systems, like a traffic jam that is not caused by one car, the anthill, which no single ant realizes it is building, like a bee, which does not know how its fellow bees fly in this regular swarm, a town, which, irrespective of the height from which it is viewed, reveals the same regular organization, a membrane that clots and a molecule that reproduces itself.

IT and all these subjects share these successive levels of abstraction. They can all offer new, sometimes astonishingly rich, functions, viewed from above, although nothing prepares them for doing so, as the level just below works by multiple interactions and iterations of very different processes and having nothing in common with these functionalities. Networks figure amongst the most commonly covered structures in publications on artificial life and give rise to a large number of emergent properties, due to the multiplicity and the connectivity of their nodes.

15.3.1 Emergence Within Networks: a Short Introduction to the Three Networks Studied at IRIDIA

Biology has revealed as many networks as anyone can imagine; genetic, cellular, neural, hormonal, immune, ecosystem, “Gaia”, trophic, (connecting prey and predators amongst animals). By adopting a unifying approach and attempting to trace relational structures common to apparently distinct realities, which is the hallmark of the sciences, the researchers at IRIDIA are convinced that the research and abstraction of these relations and the mechanisms common to these networks forms an essential part of this scientific process. In the laboratory we are attempting to gain a better understanding of the ways in which three of these networks function: neural networks, immune networks, and chemical reaction networks. The quick description of these networks that can be given is not different from that commonly understood, such as a collection of linked nodes, with the connections allowing these nodes to communicate, or more simply to influence each other (see Fig. 15.9). The current use that has been made of the word “network”, to signify social structures, networks of influence, computer networks or road or transport networks is quite similar to the way in which one can see it in biology. As Fig. 15.9 shows, when we simplify these networks in the extreme, like a collection of connected nodes, the only thing remaining is to study the way in which these elements are actually

ory and, by so doing, establishing the first major bridge between the physics of complex systems and cognitive psychology. Hopfield considered that the best way for a neural network to function as a memory is to encode each piece of information to be memorized in a fixed point of the network. The pieces of information are stored in fixed-point attractors, and when a piece of information slightly different from the memorized cell is introduced into the network, it spontaneously converges on the previously stored piece of information that is nearest to the one, which has just been introduced, because it is attracted by the attractor. It is a real emergent phenomenon and it works. Imagine a landscape with hills and valleys, with the dips in the valley corresponding to the information to be stored and the hillsides picking up all the slightly damaged versions of this information. The existence of these attracting basins explains the brain's ability to generalize and allows it to stabilize and associate a unique concept to infinity of perceived information. This is all illustrated in Fig. 15.10. Thus, primacy of coding and information processing is generally reserved for the cerebral network and although Valéry says that, "the master brain perched on its body held its mystery in its neurons and connections", the collective activation of the neurons making up the brain allows it to encode information in a structured way through the subtle redistribution of neuron activation values but also allows a second piece of information to be most naturally inferred from the first. Knowledge and association of ideas – the building blocks of human cognition – are to be found there, buried in the collective working of these neural webs.

However the connection matrix must respect two major constraints to prevent the network from exhibiting no more than fixed-point dynamics: the neurons must not possess autoconnections and the connections must be symmetrical. These constraints have no real biological parallel, but in the last analysis it has little importance to physicists who initially choose to prioritize stability and convergence over biological faithfulness. Unfortunately we have shown that by limiting ourselves to fixed points, our capacity for encoding information is very low. At IRIDIA, inspired by the large amount of neurological data produced by the electroencephalogram and other cerebral measuring devices, which really show oscillatory and chaotic dynamics in the brain, we have taken up Hopfield's developments and after freeing ourselves of the constraints imposed on connections, have also chosen to store information this time in the cyclic attractors. A cyclic attractor is defined by a dynamic behavior of the neurons oscillating between several values (Bersini 1998; Molter and Bersini 2003). In doing this, the encoding capacity is multiplied to an extraordinary degree. Furthermore, the more we seek to use cyclic attractors for storing information, the more we force the network to adopt a spontaneously chaotic dynamic, when it is not forced into one of a specific cyclic attractor (Fig. 15.11 illustrates this result). Furthermore, many neuroscience researchers are interested in cyclic systems and their possible synchronization to link up different parts of one pattern of information. Very roughly, in one still life painting,

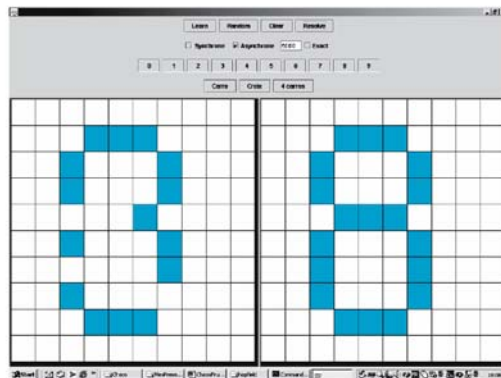
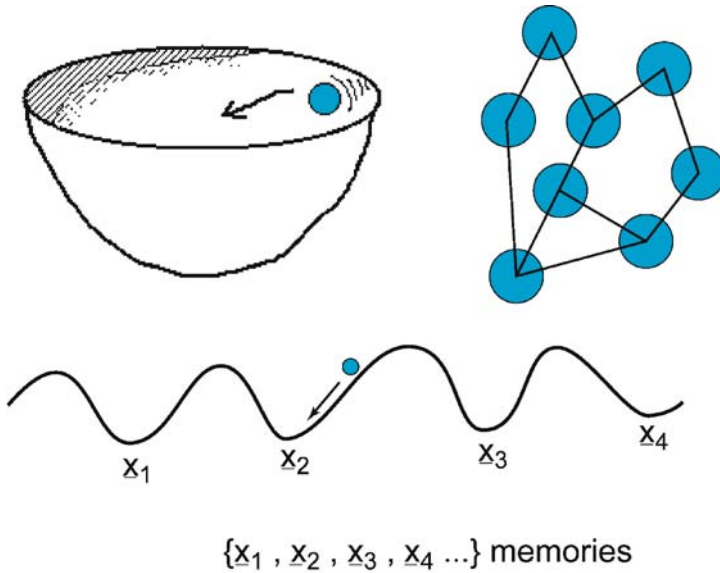


Fig. 15.10. Hopfield networks working with fixed-point attractors. In this network, connections are symmetric and the neurons binary. The memories are fixed-point attractors of the dynamics of the network. The network that has learned the figure eight finds it back on the *right*, departing from that same figure but in a noisy version. On the *left side*, one can see the initial condition of the network neurons activation, on the *right side*, the final attractor (the one that has been learned and coded in the synapses of the neural net)

the banana appears yellow and the apple green, because the neuron coding the yellow located somewhere in the brain, will oscillate in phase with the one coding the shape of the banana, but far away from the preceding one; the neuron coding the green will oscillate in phase with the neuron coding the sphere.

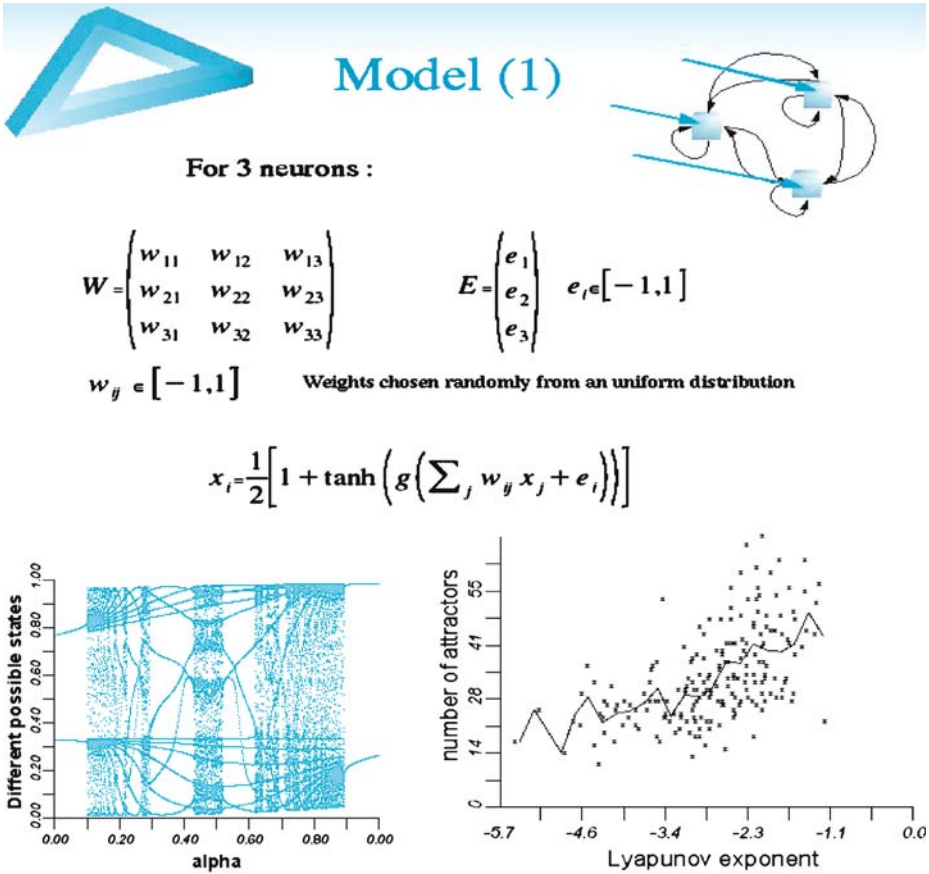


Fig. 15.11. On the *top* part, you can see a three-neuron network and the mathematical equations to which the neurons obey. The *bottom left* shows a bifurcation diagram obtained by varying the inputs to the network. It is easy to see the learned attractors and, between them, the chaotic attractors. On the *bottom right*, the plot shows the calculation of the greatest Lyapunov exponent (indices of chaos) as a function of the number of learned attractors. Clearly this function is monotonous. The more the networks tend to store information, the more chaos prevails as the spontaneous regime in between the learned attractors

Kauffman (1995), certainly one of the most influential historical researchers of artificial life was interested in the behavior of so-called Boolean networks, where the activation functions for the units is reduced to Boolean functions (only giving 0 or 1 on entry or exit, like the functions or logic gates illustrated in Fig. 15.7). He saw in this a relatively faithful response to the interactions that could exist in a collection of genes and thus became interested in the dynamic behavior of this very simplified vision of genetic interaction networks. Thirty years ago, Kauffman discovered that when a network is connected in a certain

way, for example, very weakly, each gene with two others at most, the number of cyclic attractors (which are very short on average) is considerably reduced. This is despite a very high potential number, for fixed points around two to the power of the number of units (for a network of 1000 genes this would make 2^{1000} potential fixed points, a gigantic value, as expected). The biologist was surprised to discover that both the length of the cycles and the number of cycles equals the square root of the number of units, around 30 cyclic attractors, with an average length of 30 per 1000 genes. The first reason for this low number is the emergence of a large set of genes, which do not oscillate, as if crystallizing the network and thus, common to all cyclic attractors; as for these attractors, they only concern subparts detached from the network. Kauffman's stroke of luck is to have seen that the relationship of the number of cyclic attractors in the genetic network of the square root of the number of genes is an approximate response to the relationship between the number of cellular types known to result from genetic activity and the number of genes making up these cells. All the cells of a biological organism possess the same genotype, in Kaufmann's terms, the same genetic Boolean network. These cells may however be very different, the mechanism singling them out being known as cellular differentiation: immune cells, liver cells, neurons and others. The number of cell types is estimated at being around 300, in a human being. This number rises with the complexity of the organ under consideration; it is one or two in a bacterium, three in yeast, around thirteen to fifteen in elementary organisms such as hydra, and sixty in a fly. By a happy coincidence, this number is very close to the square root of 80 000, the number of genes estimated to characterize a human being, though this is still under discussion. The number of cell types of more elementary organs seems to respect this same dependence on the square root of the number of genes.

While researchers at IRIDIA are delighted to discover that a network of three neurons allows hundreds of attractors, Kaufmann is delighted for the opposite reason: that a network of 100 genes only authorizes 10 attractors. It is an astonishing fact that although both are working on the same type of network, IRIDIA is pursuing a very different type of animals. If the laboratory of IRIDIA increases the number of attractors tenfold, it is thanks to the variation in the entries to the network, which are absent from Kaufmann's networks. It goes without saying that comparing cell types to simple attractors of this genetic network is an extreme simplification of the mechanism of cell differentiation from genetic activity. Here however it is a result of the voluntarily reductive and representative process particular to artificial life. We know this mechanism to be extraordinarily more sophisticated and still quite mysterious to most biologists: DNA, RNA, enzymes, proteins making up the pieces and essential agents.

15.3.1.2 Immune Networks

Since the 1970s and 1980s, many immunologists (the most celebrated being the 1974 winner of the Nobel Prize for Medicine, Niels Jerne) have realized that

the antibodies responsible for the adaptive defense of the immune system, by attaching themselves to any intruder, show the surprising peculiarity to behave among themselves such as with the intruder, making the very notion of intruders much more delicate than it seems. It is no longer the case of simply detecting an invader and rejecting anything that is recognized as such by an antibody, because then they would reject each other. The antibodies are no longer only capable of detecting intruders, but they detect each other, thus forming a symmetric affinity network (as illustrated in Fig. 15.12, a connection exists between two antibodies when they recognize each other). As Valera and others have extensively described (Valera and Coutinho 1991; Stewart 1994), the existence of

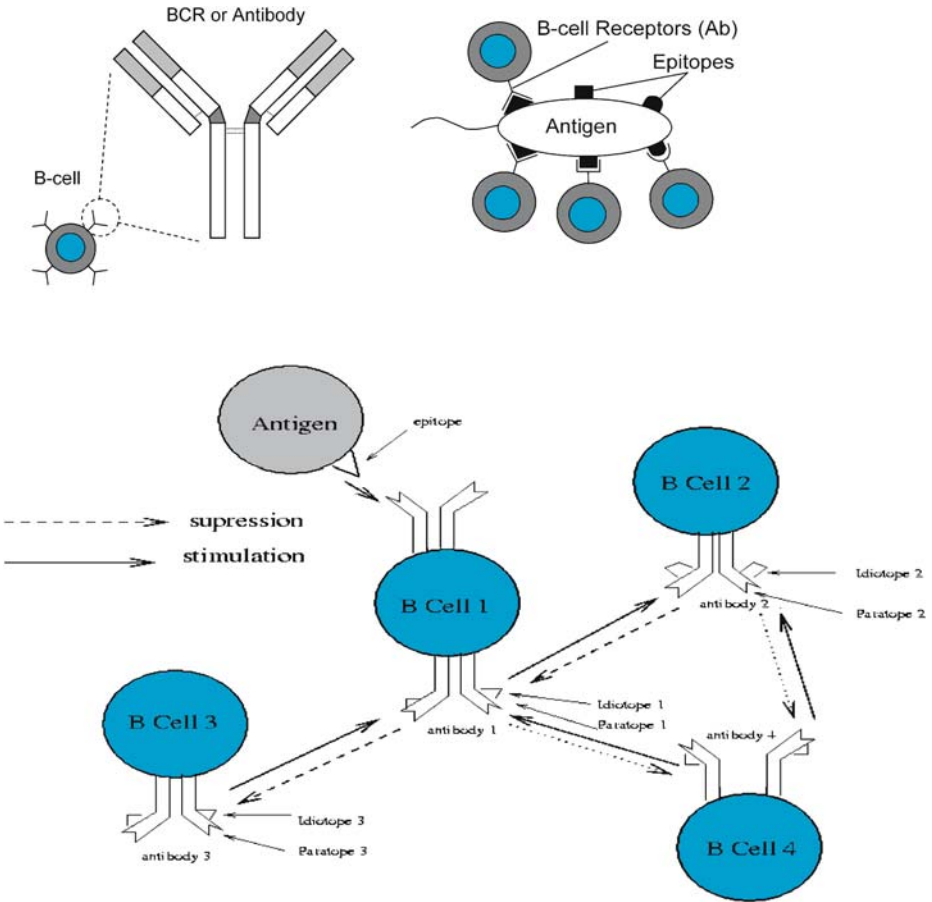


Fig. 15.12. In the classical scheme of immunology (*top*), the B cells produce antibodies, which recognize the antigen, and by binding to it, neutralize it. The idiotypic network hypothesis (*bottom*), allows antibodies to recognize and bind among themselves, forming as such a network of mutual affinity

this immune network again throws some light on the vexed question of self and non-self, the very foundation of immunology, ambiguous and highly debated as it is. Where does the self begin and what is this distinction based on? Is it linked to the characteristics of the intruder or does the system decide for itself, depending on its current *mood*. In other words, can we label as self and non-self anything impacting the immune system based on easily discernable characteristics, or is this distinction the result of a process of development where the network progressively builds up its barrier by organizing itself and building itself up, partly in response to these impacts. In this case nothing from outside the system would end up as self or non-self, but would come from inside the system. The barrier is built up gradually, its shape resulting constantly from the internal activity and situation of the system.

A very rough analogy may be drawn between a night club bouncer and the selection he makes at the entrance. His choices are partly dictated by the attractiveness and the physical or social characteristics of those who turn up at the entrance. This is not entirely the case however; the decision is also dictated by what is happening inside the club, i.e., the bouncer cannot let anyone in if the club is packed full. He can decide to give preference to women if there are not enough inside, or a member of a particular social group if he wants to re-establish a certain kind of balance. By analogy then, the decision to admit a visitor is not made solely on the visitor's characteristics, but also depends on the state of the network at a given moment. No microbe or antigen is born self or non-self, but becomes so and may even alternate between these two states, depending on the conditions within the system that it wishes to enter. The most critical dysfunctions of the system and the diseases which they bring would no longer only be linked to the problem of permeability, as the system has been attacked by an undesirable element, but rather by an internal disruption to the internal state, possibly, but not necessarily due to this undesirable element. This new vision involves a more advanced study of this internal regulation, as well as the homeostatic condition for which it would seem responsible.

The results of the simulation presented in Fig. 15.13 show the behavior of an immune system, first of all in the absence of a network and then in the presence of a network of antibodies. In these simulations (Bersini 2003) an antibody is characterized by its position on a map and is connected to the antibodies placed symmetrically in relationship to it. In the first case, the antibodies disappear immediately, provided that they are not stimulated by the antigen intruder. When the antigen appears, it stimulates the antibodies, which recognize it (the process of recognition is symmetrical). They then latch on to it and gradually make it disappear, to revert to the original situation. By playing on the speed at which the antibodies disappear in turn, the simulation is able to reproduce an inertial type memorization phenomenon, well known to immunologists (the next time this intruder appears, the antibodies could react more quickly). The second simulation is more interesting and more original, as the antibodies now

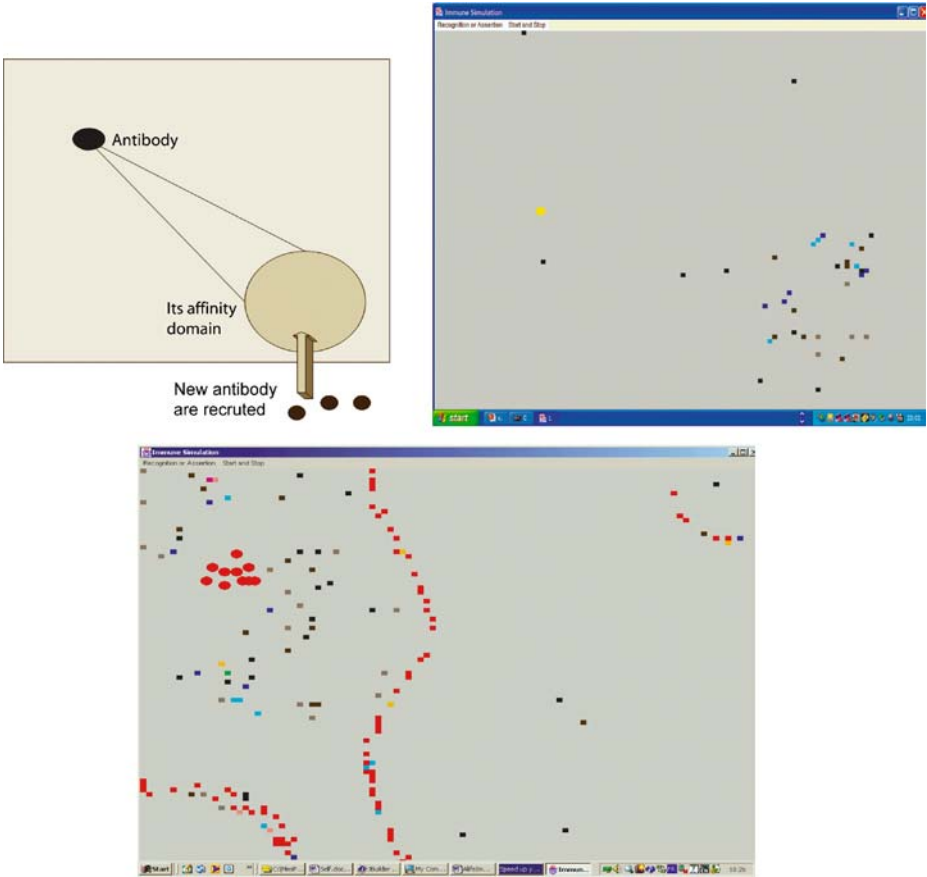


Fig. 15.13. An antibody is characterized by its situation in a 2D map. It exerts an affinity on the antibodies situated symmetrically to it. In the *top right* simulation, illustrating classical immunology when an antigen is recruited (*left*), it exerts on the zone symmetrically situated to it a field of affinity that allows to recruit antibodies able to bind to it and to neutralize it. Once the network appears (*bottom*), the antibodies mutually sustain themselves and form in the simulation lines, which separate the space in zones, which have become tolerant (antigens on the *left* indicated by red balls are tolerated) and zones, which have become reactive (on the *right* no antigen can survive). These two zones are the result of a genesis and a stabilization of a self-maintaining repertoire of antibodies

stimulate each other and cause a network that has mutual affinity and is thus self-maintained. On the results of the simulation, we can see the lines of self-maintained antibodies gradually appearing and separating the space into two zones; one in which the antigens are tolerated and one in which they are not. When the antigen attempts to enter the system, depending on the kind (the place

where it appears) and depending on the state of the network (i.e., the result of its genesis, here, the position of the lines) it will be accepted or rejected. No antigen will any longer be systematically accepted or rejected, as in the first simulation. The network divides the antigens into one category that it chooses to integrate (the self) and another that it rejects (the non-self), but this division does not exist a priori and is learned or discovered by the network itself as it develops.

15.3.1.3 Chemical Reaction Networks

IRIDIA gives priority to the study of chemical reaction networks, because, as we have been discussing since the first chapter, like the artificial life community, it considers the networks to be the main protagonists in the appearance of life, quite simply opening the doors of immortality. These chemical reaction networks where the elements are the molecules participating in the reactions and the connections are the reactions linking the reacting molecules to the molecules produced are generally characterized by fixed-point dynamics, the chemical balances during which the producers and the products mutually support each other. The attractors in which these networks fix themselves are as dynamic – the concentrations slowly stabilize – as they are structural. The molecules participating in the network are chosen and trapped by the network as a whole. These networks are perfect examples of systems, which combine dynamics (the chemical kinetics in this case) and metadynamics, as new molecules may appear as the results of reactions, while some of the molecules in the network may disappear. Both the structure of the network and the concentration of its constituents tend to stabilize over time. Kauffman (1993, 1995) and Fontana (1992) were the forerunners in the study of the appearance and properties of these networks. Figure 15.14 illustrates the work of these two researchers, dedicated to the study of prebiotic chemistry, limiting the reactions studied to polymerization, such as: $aa + bb \rightarrow aabb$ or inversely, de-polymerization: $aabb \rightarrow ab + aa$.

Kauffman (1995) showed that if the probability that a reaction takes place is affected by the presence of a catalyst, which is itself produced by the network (autocatalytic network), a phenomenon of percolation or phase transition, characteristic of this type of simulation, is produced. For probabilities, which are too low, the network does not pop up because the reactions are too improbable, but as soon as a threshold value is reached for this same probability, the network percolates, giving rise to multiple molecules produced by multiple reactions. Kauffman grants a privileged status to this threshold value and to the network resulting from it in the scenario of the origin of life, without really arguing the reason why such a status should exist, but passing the immense interest and enthusiasm that the phenomena of phase transition arouse amongst physicists into the world of biology. Fontana was concerned with the inevitable appearance of reaction cycles (such as that illustrated in Fig. 15.14. As described in Sect. 15.2.1, all the molecules produced by these cycles in the network, in turn

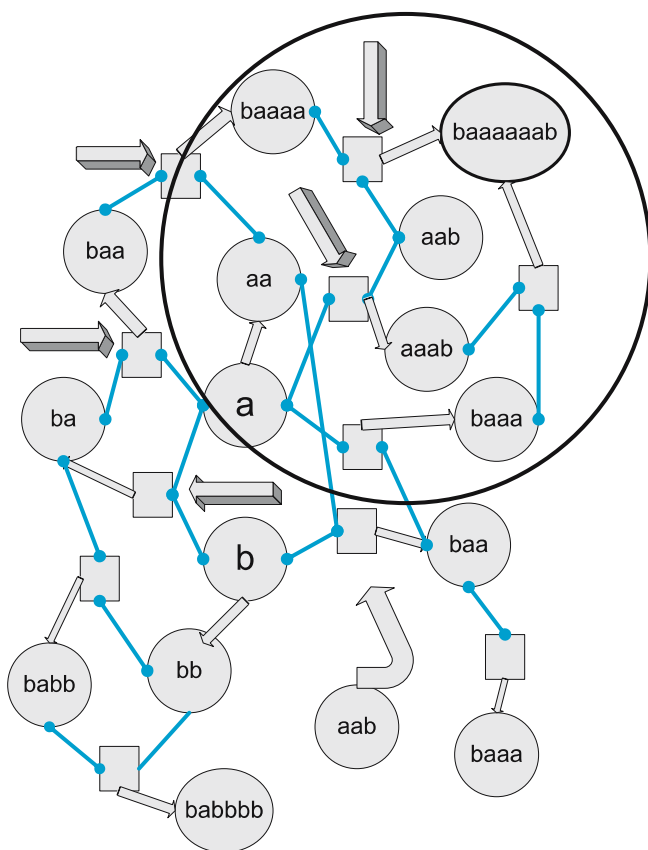


Fig. 15.14. Representation of a network of chemical reactions of polymerization ($a + b \rightarrow ab$) and de-polymerization ($ab \rightarrow a + b$) taking place in a simulated chemical reactor. Molecules are represented by *circles* and reactions by *squares*. Each reaction can be catalyzed, like the *arrows* pointing to the squares show, by a molecule of the network (giving rise to an autocatalytic network). Some molecules can appear (like the molecule *aab*) or simply disappear from the network. Reaction cycles can appear, like the one surrounded in the figure ($aa \rightarrow baaaa \rightarrow baaaaaab \rightarrow baaa \rightarrow aa$)

produce molecules of the network. Fontana is amongst those biologists, mentioned in Sect. 15.2.1, who sees these closed networks as forming a key stage in the appearance of life, due both to their stability and to the fact that they form structural and dynamic attractors for the system. They cause a stabilization and internal regulation zone in a chemical soup that is continually being crossed by a flow of matter. Fontana goes on to show how these networks are also capable of self-regeneration and self-replication. More recently, under the impulse of Plasson, Comeyras and their colleagues (Plasson et al. 2004), we have shown that homochirality, one of the most interesting and still not elucidated mysteries of

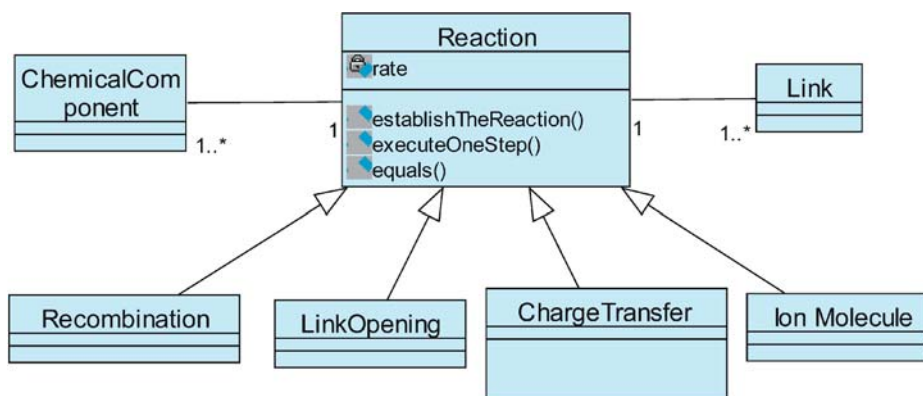


Fig. 15.15. UML class diagram of a chemical reactor with its main classes, which are: molecules, atoms, links, and many others

the origin of life (the organic molecules constituting life presenting themselves in only one spatial orientation, while they could equally well be presented in the symmetrical version) could be caused by a rupture in the symmetry, shown by one fixed point of the reactive network. The network is made up of this same type of polymerization and de-polymerization reactions. We are also studying the structural evolution of this network, the resultant topologies (Barbabasi (1999) sees connector networks in this, such as those which we will describe later, where certain molecules taking part in more reactions would play a more strategic role than others), and the emerging phenomena characterizing them.

We are programming the appearance of these reaction networks by adopting the object-oriented (OO) programming attitude. The marriage forecast between chemistry and object-oriented programming is a response to the absence of formal computer models in chemistry that could help chemists with many problems of nomenclature and classification, both of chemical types and reactive schemes. The OO allows a chemist to participate in developing computer models, particularly by using Unified Modeling Language (see Fig. 15.15), the new modeling language recognized and standardized throughout the whole computer community. The OO marks such progress towards reconciling a program and common sense, that it would be a pity not to share its benefits with other disciplines such as chemistry. It is now possible that attempts at modeling, such as those which have been introduced here, could, in addition to all the other tools that computational chemistry provides, help chemists to make more accurate predictions of certain chemical situations, or, more simply, to highlight and illustrate experimentally certain qualitative principles which they know. The OO simulator introduced here aims to reproduce a chemical reactor and the reaction network that emerges from it. Starting with some initial molecules and some initial reaction schemes, the simulator allows us to follow the appearance of new molecules, the reactions, which they cause and maintain as well as the development of their

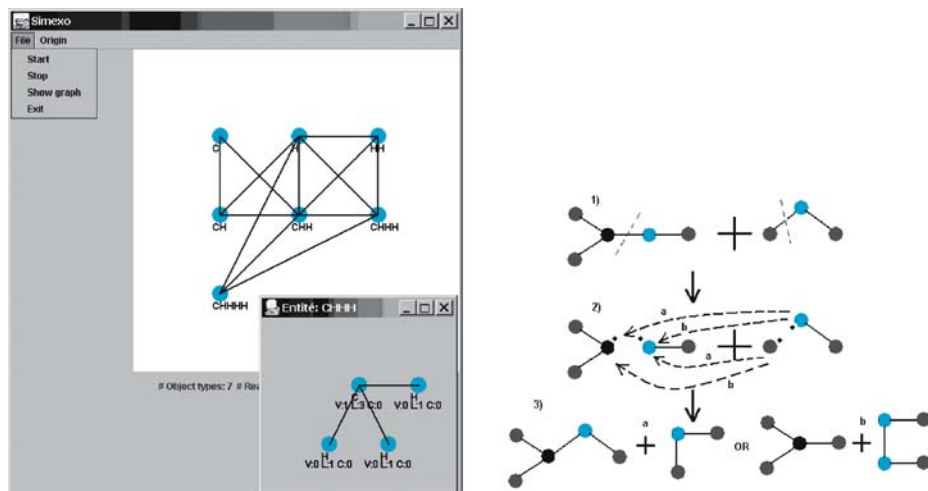


Fig. 15.16. One view (*left*) of our simulated chemical reactor and of a reaction network obtained from initial molecules containing carbon and hydrogen. Any node of the network represents a molecule and a link between two molecules exists when the first one gives the second one by a chemical reaction. One can see the molecules appear: CH, CH₂, CH₃ and CH₄ and the graph of the CH₄ molecule. One of the reaction classes is illustrated (*right*). It is a substitution reaction where parts of the reactant molecules are exchanged to produce a couple of new ones

concentration over a period of time. In terms of computer modeling, the two principal entities, which we find are the molecules participating in the reactions and the reactions themselves. The molecules are coded like graphs. They are made up of atoms and links, which open, close or break during the reactions. One molecule may be made from a reaction breaking an existing molecule or from the combination of two other molecules. Around twenty classes of reactions are coded into the simulator. When a reaction of the type $A + B \rightarrow C + D$ takes place, the concentration of the molecules A and B will decrease, while the concentration of the molecules C and D will grow by the same amount. In our simulator, the speed of the reaction K_{ABCD} , in conformity with the kinetic theory of chemistry, is calculated as follows: $K_{ABCD} = \alpha \exp(-E_{ABCD}/\beta T)$, α and β are two parameters, which will be the attributes of the reaction objects. Their significance is linked to the orientation and the shape of the molecules entering the reaction. T is the temperature at which the reaction takes place. In our simulator, the values of these parameters will be fixed arbitrarily. E_{ABCD} is the value of the energy activating the reaction. It is this value that is responsible for the real direction of the reaction. As dependence on this value is exponential, if it is high (the reaction must therefore overcome a major energy barrier in order to take place), the reaction will have a very low yield. If, however, this value is low, the reaction will take place more easily and more rapidly.

The principal attributes of the reaction class, which covers all the reactions, come from its relations with the chemical components class, (each reaction is associated with four of its components at most, neutral or ionic, molecular or atomic) and the link class, (each reaction is associated with four links at most, which break into the original molecules and re-establish themselves on arrival). This class also has two real attributes, as it stores the speeds of reactions in both directions. The reaction class is an abstract superclass, giving rise to a number of concrete subclasses making up all the subclasses in our reactor. In the simulator, three vectors of objects are important. The first vector contains all the chemical components existing in the reactor. The second vector contains only those chemical components, which have appeared at the last stage of the simulation. This separation between the two vectors exists so that the next reactions take place between all the elements of the first vector and those of the second only. The third vector contains all the reaction objects. At each stage of the simulation, new reactions are created by testing all the reactive schemes possible between the components of the first vector and those of the second. The final step is to carry out a simulation on all the reaction objects, which already exist. If care is not taken, a combinatorial explosion of new molecules takes place. In the simulator we propose several methods of overcoming this, such as allowing only sufficiently concentrated molecules to take part in the reactions. The reactions are limited to breaking only the weakest links. The result of one simulation is shown in Fig. 15.16.

15.3.2 Small Worlds

Several US physicists such as Barabasi and others have attempted to discover whether a majority of networks have a random topology and uniform connectivity structure (for an excellent review of this work, see Barabasi and Albert 1999, 2002). They have microscopically examined a very wide variety of computer, chemical, biological, social and transport networks. Still more precisely, by tracing the curve of the number of nodes in relation to the number of connections, they have established that they do not adopt the Gaussian shape, which is the sign of random phenomena, but a toboggan shape, a hyperbole or curve expressed in exponential law (in $1/x^2$ for example). The two curves are traced in Fig. 15.17. As opposed to the bell-shaped curve, this new dependence allows a small number of nodes, though more than in the case of an exponential decrease, to possess a very large number of connections. Their number remains easily lower than that of the weakly connected nodes (the top of the toboggan) but is greater than in the purely random case (the toboggan descending far more slowly than the exponential bell). These single but strategic nodes, like many crossroads on the network, have been called hubs. Physicists first discovered the presence of hubs in computer networks, both in the Internet network, which is made up of physical links and in the Web network made up of what IT specialists call hyperlinks. They were surprised to discover far more order and regularity

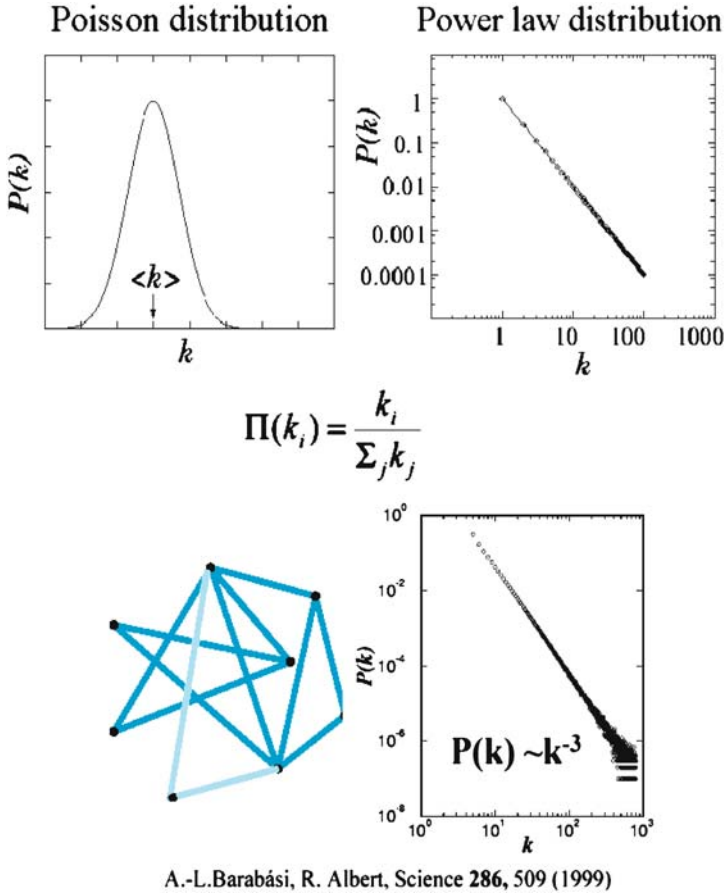


Fig. 15.17. The main results of studies conducted by Barabasi and his team. In most of the studied networks (*right*), the number of nodes as a function of the connectivity shows a power-law (a straight line in a log-log plot). This is due to the way the network builds itself, following a mechanism called preferential attachment, where new nodes, when appearing in the network, preferentially connect to nodes already more connected

than expected, despite the apparently chaotic development of these links, particularly of the Web. Since surprise is an excellent adviser in science, from this first discovery they found hubs everywhere, in practically all the networks that they explored: biological, chemical, technological, transport and up to the most social of them. Paris and Lyon are hubs on road maps, Heathrow and JFK for air traffic and the Amazon or Yahoo sites for looking at documents circulating on the Web. Physicists have attributed many virtues to this dependence shaped as a power law, such as the signature of phase transitions, and self-similar systems, also called scale invariance systems, which are not modified in any way when they are observed through either the small end or the big end of the microscope.

In the case of these networks with hubs, in order to notice them, all that has to be done is to change the nodes into macronodes, a macronode being obtained for example, by assembling four connecting nodes. A new network can be obtained, of macronodes this time, for which two macronodes would be connected if any one of the four nodes in the first network were connected with any one of the second. This new network of macronodes would exhibit the same kinds of properties as the network that they come from. These networks also exhibit the well-known characteristic of what is known as the “small world”, which allows us to interlink two nodes taken randomly, very quickly and with the aid of very few intermediary nodes. The most striking illustration is the six handshakes, linking every human being, or the diameter of the web being limited to about ten nodes (in a minimum of ten clicks on average, we can go from any webpage to any webpage).

For a particular network topology there is a particular story. The other master stroke of these physicists is to have succeeded in justifying these networks' topology in a power law by the way in which they are created and developed. This is the part of the research that interests most at IRIDIA, as they are also concentrating on the way in which these networks are made and are “unmade”, knowing quite well that nothing forces these networks to grow in a random fashion, and that the way in which these new nodes invest the network depends on the structure and connectivity already in place. This structure will be altered in turn by the addition of the new node and the fate of the next one to knock at the door will depend on the impact of the one before. Barabasi and colleagues have wondered about the origin of the hubs, which are absent from all the networks built up at random. The explanation found is as follows: they appear because the way that networks are built up, far from being random, favors them. In most of the networks examined, when a new node appears, if it is given a choice, it will prefer to connect itself to the nodes which already have the greatest number of connections, and the little set of hubs thus emerges by attracting the new nodes, which link them preferentially to the hubs. Due to this, Barabasi and colleagues have named this network development mechanism preferential attachment: for a network growing randomly, a random topology, for a network growing by preferential attachment, hubs or power law topology. It is actually quite legitimate to admit that for most networks made up of humans or designed by them (such as the Web), this method of constructing networks falls within the pattern. In a circle of friends, a new member has more chance of being introduced by the most popular member than by another. This natural amplification method consisting of putting a new member in touch with the most well-connected members of the network is too human to this extent, so we should not be astonished if we find this to be the case for most of the networks designed by man.

Like the networks recently studied by these physicists, those studied at IRIDIA are also characterized by the recruitment of new nodes over time. So in a chemical reactor, new molecules are constantly being produced as a result of reactions involving those molecules already present in the network. In the

immune network, new cells are produced by the recombination of genetic material. However, other structural developments such as a variation in intensity of the connections linking the neurons in the network of the same name are also possible in this network. As Fig. 15.18 shows we are interested in any kind of modification that these networks undergo over time. We perceive them as systems on the edge of life, re-organizing themselves around new arrivals, depending on the regulatory role that they are supposed to play within their environment. Each new arrival changes the topology of the network and thus affects the selection of the next. Furthermore the networks can do more than simply grow; nodes can equally well leave the network if they do not satisfy the criteria, which

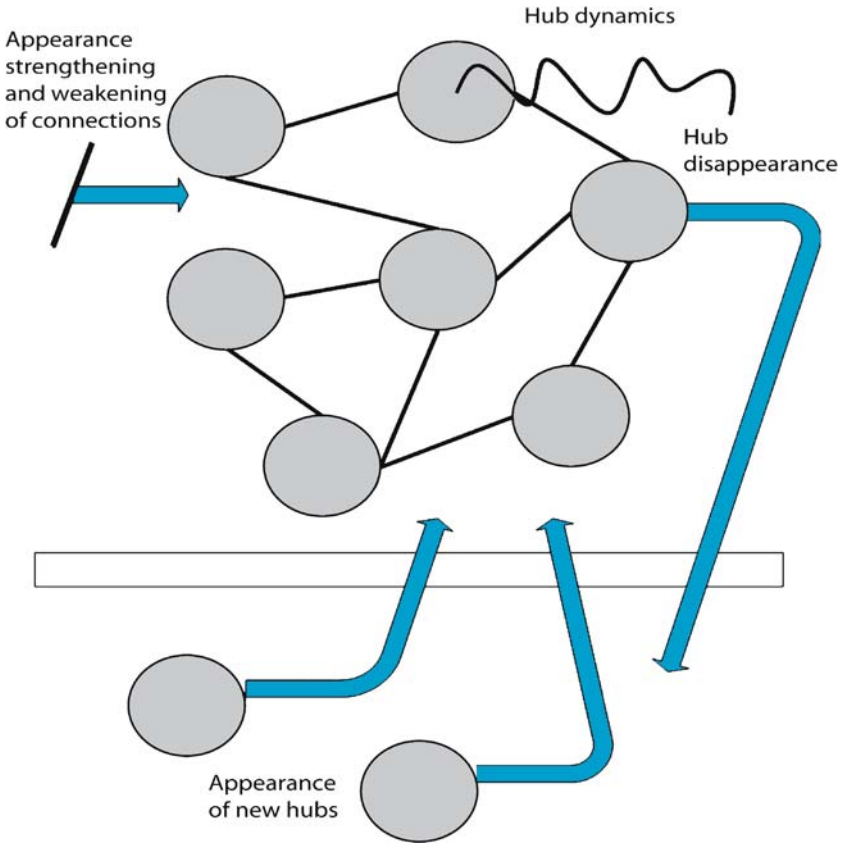
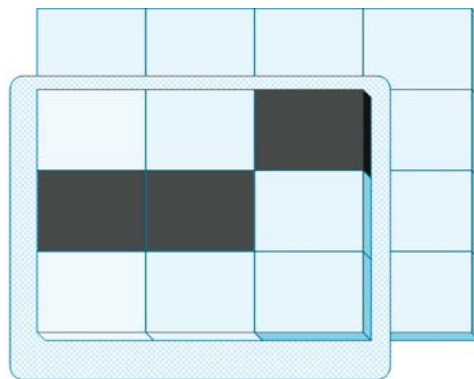


Fig. 15.18. Studies pursued at IRIDIA are concerned with the dynamical activity of the nodes together with the metadynamical evolution of these networks, by recruitment or disappearance of nodes, reinforcement or weakening of the connections. Very clearly, these two types of time evolution, dynamics and metadynamics, mutually influence each other. The appearance or disappearance of nodes will be function of the dynamical state of the existing nodes while their structural impacts will in turn affect the dynamics

allow them to stay within the networks, like molecules that leave the chemical reactors or animal species, which disappear from the ecosystems forever. The way in which the new arrivals are selected proves more subtle than a simple preferential attachment, because there comes a point where the whole network acts as the selector. Thus a node is never admitted to the network for life and it can disappear just after being recruited, if the network as a whole has decided on its fate, e.g., like a new recruit to a company who is dismissed after a six-month trial period because he is ill-adapted to his new professional environment.

15.3.3 Emergence in Cellular Automata

Cellular automata are ideal platforms for experimenting emergence phenomena, and in particular for visualizing them on-screen. Physicists and biologists are very partial to these cellular automata, as they provide an easy way of modeling situations where simple agents, distributed spatially, interact. A cellular automaton spreads out in a space like a collection of cells. Each one of these cells can be compared to a node in a network, but as opposed to networks in general and those mentioned above in particular, the spatial location of the cells has a key role to play, as each cell is connected to its neighboring cells only. Space is a central figure in the animation accompanying these automata during their simulation. The essential role played by this spreading in the programming of cellular automata provides a greater visual impact. Cellular automata are uniquely dynamic systems. As is shown in Fig. 15.19, each of the cells changes



In this cellular automata, each cell has 8 neighbours and 2 states, red or grey.
The rules are : if 2 of my neighbours are red, then I become grey, if 3 of my neighbours are grey I become grey

Fig. 15.19. Example of a 2D cellular automata. Each cell could for instance be characterized by eight neighbors and two states. The rules of evolution of the cells depend on the state of their neighbors

state as a simple function of the state of the neighboring cells (a few rules generally suffice).

15.3.3.1 Game of Life

For example, in a 2D cellular automata called the game of life, each cell has a right to only two states (such as alive or dead or 0 or 1). The way that it evolves depends on its eight neighbors and on its current state. The three rules of evolution are:

1. If the current state of the cell is 1 and two or three of the eight neighboring cells are at 1, the state remains at 1.
2. If the current state of the cell is 0 and exactly three of the eight neighboring cells are at 1, the state changes to 1.
3. In all other cases the new state will be 0.

The simulation must be carried out synchronously, i.e., all the cells must update at the same time. Any other form of update (for example asynchronous, one cell at a time chosen by chance) generally devalues the emerging phenomena, as it tends to stabilize and freeze the collective behavior of the cells (Bersini and Detours 1994). Some cellular automata can make very pretty patterns, as illustrated in Fig. 15.20 and remind us of some patterns characteristic of biological organisms.

In the game of life, the collective behavior of all these cells over a period of time allows a little motif called a glider, made up of five cells in state 1, to move around, as shown in Fig. 15.21. The glider's movement comes from simply carrying out three rules of the cellular automata, as none of the movements is coded or shows through in the elementary rules of the working of the automata. It is emergent, because it is undeniably new, perceived as such by a human being, due to the interaction between the cells, whereas the cells' updating rules did not predict anything. In the game of life, the behavior of the automata is so complex, despite the three elementary rules sufficient to make it function, that hundreds of pages are not enough to describe the immense variety of small-scale phenomena taking place before our very eyes, depending on the initial conditions

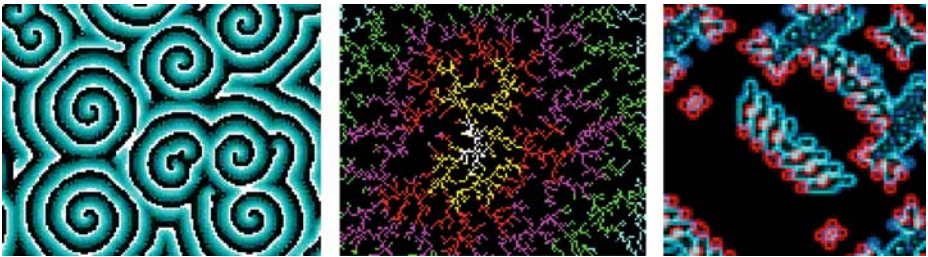


Fig. 15.20. Nice patterns obtained by the simulation of 2D cellular automata

(Poundstone 1985). It is easier to construe this new dialogue between the simple and the complex, only made possible by computer simulation, as it would require an encyclopedia to describe the behavior of the whole, whereas a few lines are enough to describe the behavior of the parts.

15.3.3.2 Morphogenesis and Self-Replication

An astonishing historical fact regarding artificial life and, more particularly about the study of cellular automata is the importance of the role played by two scientists in the forefront of the field, who we have already been mentioned and to whom we owe the discovery of the fundamental principles governing the way in which computers work: Turing and von Neumann. This is not due so much to a happy coincidence, as to the fact that the mechanisms underlying the way in which computers function are closely related with biological phenomena, and the fact that these two geniuses realized this. We may legitimately feel that although its official birth was at the first US conference in 1987, the process of using the computer for a better understanding of the mechanisms that give life to inert matter and conversely perfecting computer processes based on this understanding is as old as the first steps in computing. Thus Turing, the designer of the universal machine, also asked questions about the origin of shapes in the animal world, for example the characteristic motifs on animal skin (zebras, giraffes, ladybirds) or shellfish. In an extremely synthetic way, (see Meinhardt 1998, for a wider view of this marvelous visual field of study), a cellular automaton (such as the example in Fig. 15.22) will be characterized by the appearance of zebra stripes or other varied and alternating motifs if the state of a cell is influenced positively by those which are nearby, but, in contrast, negatively by those which are further away. We can understand intuitively that updating rules of this kind, exciting anything that is nearby but inhibiting anything that is far away, will create bands of cells at 1 (or active) surrounded by bands of cells at 0 (or inactive), reflecting quite faithfully the skin of zebra or giraffes. This discovery of

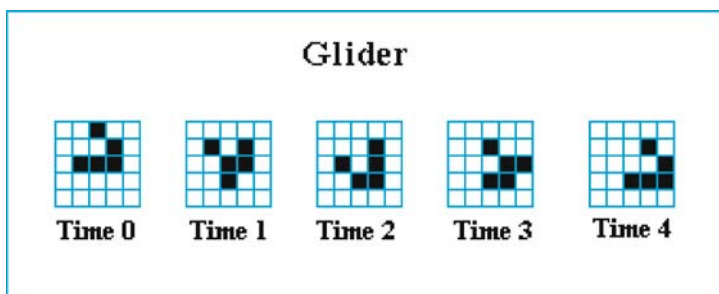


Fig. 15.21. In the game of life, a little motif called glider moves around as a result of the synchronous updating of the three rules presented above

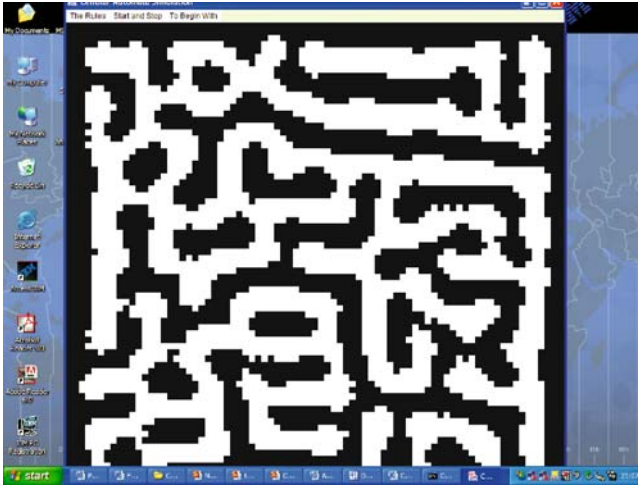


Fig. 15.22. Simulation of a 2D cellular automata in which the updating rules are so that each cell influences positively its neighbors and negatively the cells far away. The emerging motif is alternatively composed of *black lines* (active cells) and *white ones* (inactive cells)

Turing set off a whole new field of theoretical biology dedicated to animal skins and more generally the morphogenesis of biological organisms.

As we have seen in Sect. 15.2.4, after the universality of Turing's machine was made more concrete and tangible through step-by-step programs, which could be loaded into the memory of the computer like the data that they process, von Neumann, following in Turing's footsteps, tackled a biological mystery at the time: self-replication. Ten years prior to the discovery of DNA, he considered the basic structural and functional ingredients, which each machine capable of self-replication should contain, based on the example of the elementary cell. Von Neumann gave a cellular automata solution of the problem described in Sect. 15.2.4. Each cell possessed five neighbors and 29 states, and around 200 000 cells were necessary for the phenomenon of self-replication to take place. Many years later, Chris Langton, (1984) the organizer of the first conference on artificial life, would propose an extremely simplified version of this (eight states, but 219 rules remain necessary), although it still follows the pattern mapped out by von Neumann. This automaton, shown in Fig. 15.23 incessantly reproduces a little motif shaped as a loop. Here again, self-replication seems to be an emergent property as only we, the observers, are able to see it reproducing itself and none of this kind of phenomenon is covered in the updating rules. In a very recent and book Wolfram, one of the designers of these cellular automata, has made them the actors of a new kind of science (Wolfram 2002). With all options for new thinking and new choices still open, Wolfram seeks to portray himself as the guru or spiritual leader.



Self-Reproducing CA

Fig. 15.23. The self-replicating cellular automata (Langton 1984)

15.3.3.3 Autopoiesis

In the examples of cellular automata examined previously, the cells do not move and are limited to changing state. In the wake of Varela et al. (1974), some researchers interested in the methods of cellular individuation discussed in Sect. 15.2.3 have added cellular movement and, in particular, the faculty of cells to be surrounded by a membrane: the only thing capable of conferring the status of an individual on them. This is the case for McMullin, who is continuing the work of Varela (McMullin and Varela 1997) and attempting, as shown in Fig. 15.24, to create a minimal simulation of autopoiesis. This model covers three types of particles capable of moving round a 2D surface: substrates, catalysts and links. The working and updating rules of this cellular automata are as follows:

- If both of two substrates are near a catalyst, they disappear to create one single link where one of the two was located.
- If two links are near each other they link up and attach themselves to each other. Once attached these links become immobile.
- Each link is only allowed to attach itself to two other links at the most. This allows the links to form chains and to be able to make up a closed membrane.
- The substrates can diffuse through the links and their attachments, while the catalysts and the other links cannot. We can therefore understand how the product of the co-generation discussed in Sect. 15.2.3 comes about. The membranes shut in the catalysts and the links, which in turn support the membrane by being essential to its formation and regeneration if it is destroyed.
- The reactions creating the links are reversible, as the links can recreate the two original substrates (and thus cause the membrane to deteriorate), but at a lower speed. When this happens, the attachment between the links also disappears.

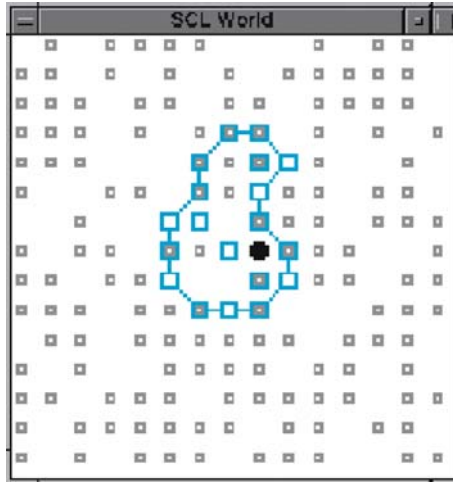


Fig. 15.24. Simulation by means of a cellular automata of the autopoietic model (originally proposed by Varela et al. 1974)

Continuous updating and execution of these rules produces minimal versions of reactive systems, physically closed and confined by means of a membrane, which is itself produced by the reactive system. For Varela and the others following him, this is an essential stage in the road to life.

15.3.4 Useful Emergence

Along with others, Bersini (2004) has recently considered the notion of emergence and has expressed some dissatisfaction about describing it as no more than new properties appearing at a higher level of observation. At this level, where the functions of different parts combine, these new phenomena are detected by a human observer capable of integrating time and space by his perception. The unsatisfactory aspect of this description is the fact that reference continually needs to be made to a human observer, an ideally surprised observer, for a phenomenon to qualify as emergent. The fact is that man is surprised to see the glider move in the game of life, to see Langton's loop reproduce itself in the self-replicating cellular automaton and the membrane close in on itself. As Fig. 15.25 illustrates, it would be more convincing if man could be replaced by another automatic system, which would calibrate its perception system itself, so that it could select certain properties observed for its own purposes. In other words, for this additional external system, the glider would have to be so useful that there would really be a need not only to detect it, but to use it for its own advantage. As scientific progress or the re-affirming of scientific pronouncements comes as a result of a strict process of "objectivation", the notion of emergence would also gain from moving away from having a human observer necessary. These

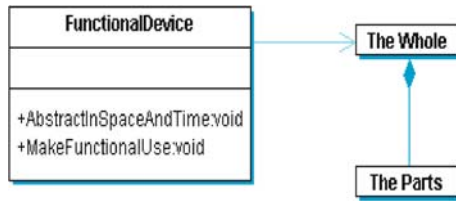


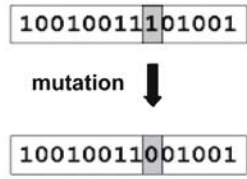
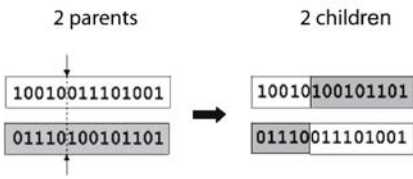
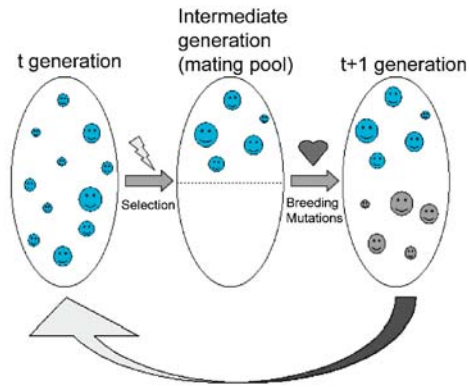
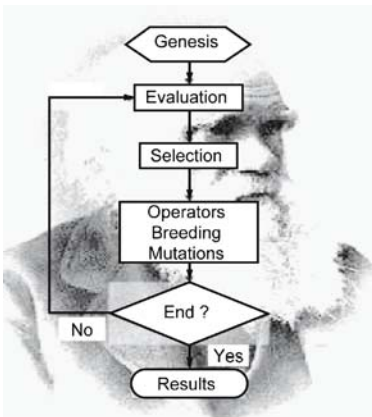
Fig. 15.25. Classically a human observer is needed to perceive and label an emergent phenomenon. Now, the formal characterization of emergence would gain from getting rid of the human and substituting with a mechanical observer able to detect and make use of this phenomenon

phenomena are characterized as emergent through observation, then selection by a system outside the one producing the emerging phenomena. As with top-down engineering systems, which each level having higher properties observed and are selected under human control so as to be translated into lower level mechanisms, for functionalities to emerge an external entity must discover, select and use these functionalities in a bottom-up way to make basic constituents of the higher levels.

15.4 Plasticity and Adaptability

Faced with this immense variety of possible types of emerging behavior, what can a biological system of any kind do to select the one or ones that benefit it? Basically, how does it become adaptable? It works by selection or reinforcement mechanisms, which allow it to isolate useful patterns of behavior amongst many. There is no need to return to the Darwinian theory justifying the improvement of species by this succession of random alterations over a period of time, linked to a mechanism selecting the best adapted. When an engineer uses these same principles, the engineer substitutes himself for nature in order to select a behavior allowing us to solve a given problem. Let us take the example of genetic algorithms (Goldberg 1994), a method of optimizations completely inspired by Darwin's theory and certainly one of the most profitable and famous uses of artificial life for engineering. When a cost function is defined associated with any proposed solution of a problem, this turns out to be a problem of optimization, and we seek to maximize the cost, for example. The running of the algorithm is illustrated in Fig. 15.26. For each iteration of the algorithm, a set of candidate solutions is proposed. Each candidate is coded through a symbolic language. This is often binary and, once decoded, will lead to a behavior, which can be evaluated. The candidate is evaluated by calculating the cost function, sometimes called fitness, recalling Darwin's theory. Three genetically inspired operations follow (hence the name of this algorithm), which allow the candidates to progress to solutions, hoping that this progression will be positive: selection, change and recombination.

During the selection, in the following population, we retain only one subset obtained from the best candidates coming from the preceding population. This selection not only preserves the best candidate, at the risk of converging too quickly on a local (and not global) maximum, but picks the candidates in terms of probability, still favoring those which show the greatest degree of fitness. This intermediate population retains only the most promising solutions from the preceding population, including some outsiders, which allow the research to continue over the whole area. The manipulations to come take place amongst this elite population. In order to progress in the solutions area, i.e., generate new candidates, we have to leave this population and make some alterations to the existing candidates. In order to do this, two operations come into play: mutation and recombination. During mutation some of these candidates are altered at random (for example, in one of the randomly chosen candidates, the value of one bit, also picked at random, is inverted); the hope being that an erratic search of the good candidates allows us to discover better, or on the contrary that if the candidate is lost in a blind alley, this shake-up can deliver it. During the recombination, (here, sex intrudes into the algorithm), two candidates are crossed in order to create two new ones, obtained from one subpart of one



Recombination

Fig. 15.26. The genetic algorithms (above): the recurrent sequence of selection → mutation and recombination. Examples of the two classical genetic operations (below): mutation and recombination applied in solutions coded as a binary chain

and from one subpart of another. It is then hoped that the new candidate, the offspring, could thus inherit the things that make up the quality of its two parents, joint qualities, so that a better solution could be found. Recombination is the ancestral practice of breeders, who inspired Darwin to develop the theory despite living a long time before him. This recombination is common currency in many optimizing procedures, for example the selector of the national football team, who combines the regional teams in the hope of finding the ideal hybrid. At first, many forms of creativity amount to the recombination of known and memorized motifs, as jazz improvisers know well.

These are the same genetic algorithms used by Dawkins when developing biomorphs (Fig. 15.27). They also form the basis of many ecosystem simulations, where they are found at the birth of new species trying to improve their adaptive value. It should also be stressed that another element in Dawkins' program is that when it is finally evaluated, the phenotype is not directly obtained from the genotype, as would be the case for classical optimization in a real or combinatorial space. In this work, these types of insects are the result of a recursive program that is carried out starting from a given genotype to give a phenotype. A great semantic distance is maintained between these genotypes and phenotypes, which reflect the long process of cell construction from the genetic code. In the same way, a very prized derivative of these algorithms is genetic programming (Koza 1992; illustrated in Fig. 15.28) where the individuals to be optimized are programs. This is carried out after they are executed and fitness is attributed. They are combined with their peers in the hope of producing a better version of these programs. It is a genetic version of every computer engineer's fantasy; that of developing automatic code generation and simply announcing the goal that is pursued. All these computer experiments show very rapid solution of these astonishingly complex problems after very few iterations over the scale of evolutionary time. Our directed chemistry that was the subject of the last section can equally be perceived as the simulation of a form of an ecosystem, as two types of dynamics are happening at the same time: development over time of already existing

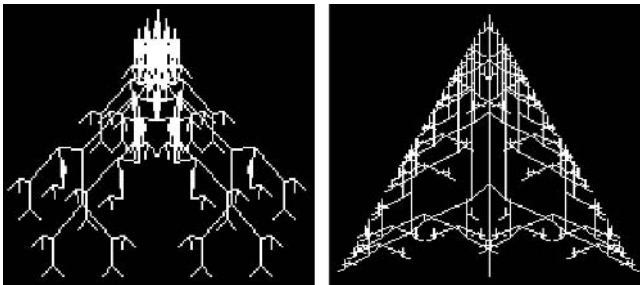
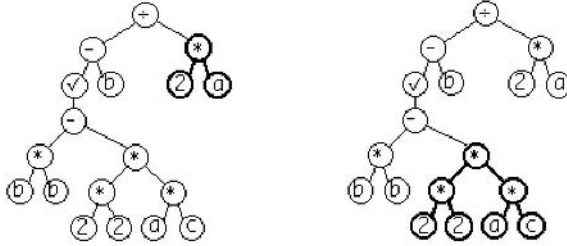


Fig. 15.27. Two biomorphs of Dawkins (1989). These figures are produced by recursive algorithms executed from a given genetic code that evolves by means of genetic algorithms

Crossover Operation with Identical Parents

Parents



Children

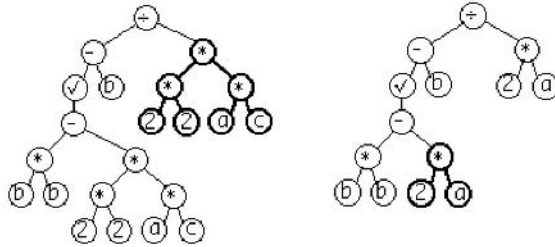


Fig. 15.28. Genetic programming. Two programs, which, once executed will give a certain result, are recombined in the hope to find still a better program

species and the appearance of new species by the recombination or alteration of the chemical species present (due to the reactions).

For the engineer, these elementary but innumerable produced mechanisms, can reveal high-quality solutions to problems regarded as insoluble (e.g., complete Non-deterministic Polynomial-time (NP) problems like the famous traveling salesman problem, where the time required to find the solution depends exponentially on the number of towns), which plague the world of business: payments, planning, timetables, re-grouping problems and stock management. Even processes of artistic creation can be affected by this original method of production, suggesting new musical, pictorial or architectural forms, by recombining those which have proved most promising up to the present. By its amazing brute force and this little touch of biology, the computer can tackle applications supposed to be unsolvable. Through these algorithms the engineer accepts this loss of control mentioned in the introduction. In fact the engineer is limited to guiding and directing the research process, but is excluded from what is strictly called the production of solutions. There are no mental means of accompanying

RL is Learning from Interaction

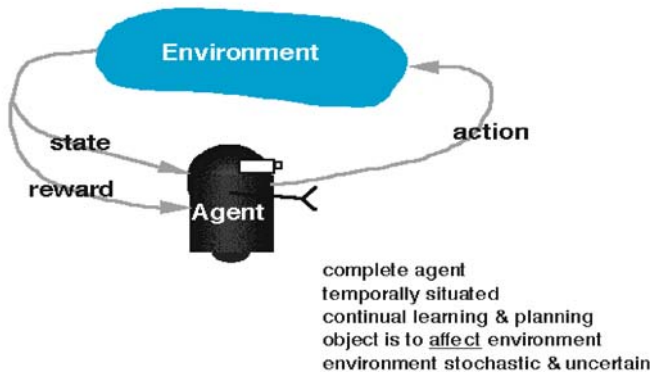


Fig. 15.29. Reinforcement learning. An agent endowed with sensory-motor capacities circulates through its environment and improves its behavior by receiving a feedback in response to its actions, rewards or punishments. It aims at improving its sensory-motor associations by maximizing the sum of rewards over time

the computer during its unrestricted generation of possible solutions and it is content to accept reasonable proposals and to sort out the chaff from the wheat. The whole problem comes down to encoding the possible solutions properly, a correct definition of fitness (as this notion is even extremely delicate to define in Darwin's theory) and an ingenious adjustment of the genetic operators.

If the genetic algorithms make up a biological adaptability mechanism working over multiple generations, then reinforcement learning (Sutton and Barto 1998) happens over a far shorter timescale, as it is a neuron-type learning process, taking place within the life of the biological system. In this case the parameters, link or agent (this often means synapses linking two neurons between them), having contributed positively to obtaining a correct solution are positively reinforced whereas, in contrast, we inhibit the parameters, which have contributed positively to obtaining an incorrect solution. Many robots with motor-sensory capacity, which are wandering through a real physical environment (see Fig. 15.29) adjust the way they work to fit in with this pattern so as to discover which one of the possible parametric combinations leads to the right port (for example when they have a number of articulations to calibrate). In these two forms of apprenticeship we find the same necessary balance between purely erratic movements, when looking for solution zones to examine more carefully and a far more instructed and determined exploitation of these zones. To quote the title of the celebrated work, they are the algorithms, which proceed to the crossroads of chance and necessity. This determinist approach is always made by favoring the best proposals and detaching oneself from the bad, to generate the next. Finally, we have to be happy about this big loop,

where biology comes back to biology; we discover that artificial life gives rise to a set of tools that now has a vast field of application within biology itself, such as the neural networks for reading and interpreting the genetic microarrays, or algorithms of the same name (genetic) for the prediction of the 3D conformation of proteins.

15.5 Environmental Autonomy and Significant Integration

Parallelism, functional emergence and adaptability are the conditions necessary to allow these new biologically inspired artifacts to emerge. We are jumping straight into the robotics branch of artificial life, with its most popular mascots represented in Fig. 15.30. Here again we find the leaders of one of the most ferociously critical campaigns ever directed against artificial intelligence, who aspire to see artificial life as the product of an irreversible secession. They accuse artificial intelligence of neglecting the embodiment or the genesis of their supposedly expert or intelligent systems by making them reason against a symbolic or representational background provided by the user. However, a robot in a real situation has to operationally and independently construct its own symbolization of the world, i.e., a useful representation allowing it to move within the world and to remain viable. To paraphrase Brooks (1990), the uncontested leader of this new robot world: it must be plunged into a real world, a world that remains the best representation that we can make out of it. Imagine an expert system capable of medical diagnosis, the jewel in the crown of artificial intelligence, in the doctor's surgery – its natural environment. The patient arrives. Without the doctor putting the computerized expert in place, without the symbolic background from which the reasoning operate installed, for example a description of the patient and their symptoms, the expert system will turn out to be as efficient as a pot of jam.

Today some vehicles can move without a driver on a US freeway. Here again, it is perfectly possible to predict all possible situations and to supply this pre-symbolization to automatism. But what would happen to this computerized driver if he meets a body on the road, or worse still, if we ask him to leave the motorway to go into New York? The difficulty for the computer becomes how to take all potential situations into consideration automatically, from the most normal to the most unusual. For those in favor of this new biologically inspired robotics the answer lies in putting robots into real situations, by giving them multiple primitive motor-sensors, a primitive representational support that has to be supplied with basic needs (for example maintaining the power supply) and a large learning capability. When it is supplied with this basic kit, when it is in situ, the robot will provide itself with sufficient representations of its world to allow itself to find the answers necessary for maintaining its viability. This criticism is severe; it is linked to related criticisms addressed for a long time by many philosophers to artificial intelligence. Just because a computer system can manipulate (translate, simplify, etc.) stereotypical scenarios of a restaurant,



Fig. 15.30. Some famous artificial life robots

it does not entail that this robot understands a restaurant like a human being understands it. For these philosophers, syntax is not enough for semantics. Only a robot that is fed up with its diet and going for a stroll in town with some money in its pocket, seeing the luminous sign of a McDonald's for androids will really understand what a restaurant is about. The interfacing with the real world required by these robots needs a parallel information reception mechanism, because the environment subjects it to a constant bombardment of stimuli. They have to learn to organize and master this avalanche falling on their senses. They have to learn to build their own concepts, fed and stimulated by this environment, and what, in return allows them to master it. The concepts are born out of motor-sensory interactions and serve to support them.

We can understand the necessity for significant integration for the robot in its interactions with the environment. Cognitive systems act just like the minimal cell in the primitive soup, with a flow of matter crossing straight through, maintaining itself by integrating this matter to form a closed reactor network

and the membrane enclosing it. They do not passively receive a pre-determined world, but integrate it in a way that is adapted to their structure and their maintenance in the world. Here, this cognitive enclosure of beings, which think and act in their world, reflects the biological shutting off underlying the emergence of primitive living organisms at another level of description. Biological systems initially function in a closed circuit before opening themselves to the outside, and what they do with this *outside* is first and foremost a result of the concern of keeping itself alive. Finally, the motor aspect of these robots has many factors; it is emergent and needs regulation that can be adapted to it. We understand then, that on the road leading to real artificial and autonomous creatures, the three lessons of life that have just been drawn up, mutually change and influence each other. To perceive, move and maintain itself in the world that is in part the one that it will have built, the robot will be subject to many emergent and adaptive functionalities.

15.6 Conclusions

The conclusions of this chapter are directed to the three partners who already mentioned: the biologist, the engineer and the philosopher. To the first, the final result of artificial life is to bring out what the computer and the biologist share intimately: an elementary way of working at the ultimate level, but which by the brute force of parallelism and incessantly repeated iterations, can lead to the emergence of unknown and sophisticated phenomena at higher levels. These processes give rise to many different ways of being, which will be sorted by a further phase of selection aimed at a higher well being or viability of organism. The biologist must out of necessity take into account this functional hierarchy, which is the key to the ways in which computers function. What could therefore be better than reproducing these mechanisms by interposing computer programs suggested by artificial life? This is the only way of allowing the path to be cleared of any false simplicity, of operating this reading at several levels and exposing the bottom by observing the top. Many areas of biology have been invaded by computer simulations, which reproduce known phenomena qualitatively: the behavior of insect societies, evolutionary mechanisms such as punctuated equilibria or the Red Queen, extinction phenomena in exponential law (few major extinctions but many minor), calling into question self/non-self in immunology, cyclic and chaotic cerebral dynamics, phenomena relating to the origin of life such as self-maintenance, self-individuation or replication, artificial chemistry predicting the emergence of stable closed reactor networks, symmetry breaks, and many others. The qualitative aspect of these simulations can give them new roles in the vast scientific register: use it for education, illustrate biological principles that are already understood, open up possible experiences of thought, play and replay multiple biological scenarios very quickly, titillate the imagination by on-screen representations, call into question some of the ambiguously-interpreted

but commonly-accepted facts. They should clarify, examine and test new explanatory schemes, find uses for some phenomena, which we have been able to observe up to the present. We must also realize that anyone too full of naive enthusiasm, coming up with slogans, which are too catchy by being too qualitative such as life is a phase transition phenomenon, life's origin is taken from the edge of chaos, and many others, have had the effect of rendering a disservice to this discipline, rather than anything else, particularly in the eyes of biologists. It is their responsibility to remain vigilant and sort things out: they are doing so. Without fear of contradiction, theoretical biology today is no longer isolated from the messages coming out of the artificial life community (far from it).

The second partner, the engineer, is vigorously encouraged to use the computer for what it is best at doing – the infinite possibility of trial and error. The engineer must exclude from the loop what one cannot take in, the time for the computer to propose this immense range of solutions to the problem at hand. Then, like a child playing at “hot and cold”, the ability to guide the computer will emerge, because the engineer is the only one who will know finally the nature of the problem and will be the only one capable of appreciating the quality of the solutions. It is a perfect synergy, where both participants complement each other perfectly. While the engineer must bow to the computer in terms of calculating power, this is compensated for by judgement. Genetic algorithms, ant colonies, neural networks, reinforced learning, all have enriched the engineer's toolbox. It is rare for new problems to be completely created piece by piece and we are still left with the usual data processing, with situations requiring either optimization or regulation. But these new algorithms coming from artificial life have the singular advantage of adding simplicity to performance and leaving the computer to its calculating power, and to soften the torment undergone by the engineer in the inferential progression to better solutions.

Finally, the philosopher, who has already had close dealings with the functionalist theory of artificial intelligence, attempting to convince us that the machine could, over the medium term, be confused with a conscious being in every way. Indeed even consciousness, the last bastion, would finally surrender at the foot of the computer. Its extremely efficient defense mechanism rested on the purely subjective nature of the phenomenon of consciousness. How can we honor a machine with consciousness, since the very nature of this consciousness renders it impermeable to objective analysis, the indispensable prerequisite of the chance of any possible scientific replication? We have to extract consciousness from ourselves and make it a scientific object by interposing a definition of phenomena accepted by everyone, in order to begin considering reproducing it in the machine. Consciousness however, is ill-suited to any such exercises in objectivation. What is the case for life? Can philosophy use the same defensive strategy against the functionalist theory, more modestly attributing life to these machines? Remember that for Descartes, at the time, the animal as a machine was alive but not conscious. Attributing life to a machine poses a problem of a completely different nature to a philosopher. This means finding a mechanistic

definition or list of properties, which can achieve unanimity. This can encounter various obstacles, the one of greatest consequence being that any mechanistic or operational definition, such as self-replication, maintaining a homeostatic state, metabolism, growth, the dynamics of evolution and many others are directly confronted with a computerized version of the same phenomenon (many of these mechanisms give rise to interesting philosophical questions, such as the formalization of emergence). Granting the computer access to life is not to everyone's taste. Its consequence would simply be to de-value membership.

As noted in the introduction, science studies the sky but does not define it. It studies, but does not define the universe, the sea or the weather. Why should it be different for life? For each attempt at a definition, artificial life makes a real attempt to achieve once more a computerized version in conformity with this definition. For the biologist, the sceptic or anyone unhappy with this computerized lining, the question becomes how to refine his definition, to complete it or to renounce the possibility that there is no definition that cannot be computerized. The other possibility, doubtless more logical but more difficult for many to accept, would be that life poses no problem for a computer snapshot. Bedau, philosopher and editor-in-chief of the review entitled "Artificial Life" and principal organizer of all the social activities concerned with this field of study, discussing with other philosophers his definition of life as "a continuous process of creative evolution". When people retorted that he could be talking about the Web or genetic programs, which were continuously and creatively evolving, he had, in principle, no hesitation in welcoming them into the kingdom of the living. In the end, the beneficial effect of such an attitude is to help de-sanctify the idea of life in its most primitive form, when it is the privilege of the most elementary organisms, and not, as in more evolved organisms, when it takes root and becomes indistinguishable from manifestations of consciousness; in substance this means to keep life and consciousness of life clearly separate. Thus closing with a reference to a philosophical article, which has become famous in the artificial intelligence community: if a computer cannot know what it is like to be a bat, living like one could be within its reach (Nagel 1974).

References

- Barabasi, L.-A. and R. Albert. 1999. Emergence of scaling in random networks. *Science* 286, pp. 509–512.
- Barabasi, L.-A. 2002. *Linked: the New Science of Networks*. Perseus Books, New York.
- Bersini, H. and V. Detours. 1994. Asynchrony induces stability in cellular automata based models. In: *Proceedings of the IVth Conference on Artificial Life*. MIT Press/Bradford Books, Cambridge.
- Bersini, H. 1998. The frustrated and compositional nature of chaos in small Hopfield networks. *Neural Networks* 11, pp. 1017–1025.
- Bersini, H. 2002. Self-assertion vs self-recognition. A tribute to Francisco Varela. In: *Proceedings of the 1st International Conference on Artificial Immune Systems (ICARIS 2002)*, Canterbury, UK, 2002, pp. 107–112.

- Bersini, H. 2003. Revisiting idiotypic networks. In: Proceedings of ECAL 2003, Dortmund, Germany, 14–17 September 2003, pp. 164–174.
- Bersini, H. 2004. Whatever emerges should be intrinsically useful. In: Artificial Life 9. The MIT Press, Cambridge.
- Brooks, R. 1990. Elephants don't play chess. *Journal of Robotics and Autonomous Systems* 6, pp. 3–15.
- Buchanan, M. 2002. Nexus: Small Worlds and the Groundbreaking Science of Networks. Norton, New York.
- Dawkins, R. 1989. L'horloger aveugle. Laffont, Paris.
- De Duve, C. 2002. Life Evolving. Oxford University Press, New York.
- Dorigo, M. and G. Di Caro. 1999. Ant algorithms for discrete optimisation. *Artificial Life* 5(3), pp. 137–172.
- Fontana, W. 1992. Algorithmic chemistry. In: Langton, C.G., Farmer, J.D., Rasmussen, S., and Taylor, C. (eds.) Artificial Life II: Proceedings of the SFI Studies in the Sciences of Complexity, vol. 10. Addison-Wesley, Reading, MA.
- Ganti, T. 2003. The Principles of Life. Oxford University Press, New York.
- Goldberg, D.E. 1994. Algorithmes génétiques. Addison-Wesley, Paris.
- Hopfield, J. 1982. Neural networks and physical systems with emergent collective computational properties. *Proceedings of the National Academy of Sciences of the USA* 79, pp. 2554–2588.
- Jerne, N. 1974. Towards a network theory of the immune system. *Annals of Institute Pasteur/Immunology (Paris)* 125C, pp. 373–389.
- Kauffman, S. 1993: The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press, New York.
- Kauffman, S. 1995. At Home in the Universe. The Search for the Laws of Self-Organisation and Complexity. Oxford University Press, New York.
- Koza, J. 1992. Genetic Programming. MIT Press, Cambridge.
- Langton, C.G. 1984. Self-reproduction in cellular automata. *Physica D*, 10, pp. 135–144.
- Langton, C.G. (ed.) 1989. Artificial Life I. Addison-Wesley, Reading, MA.
- Lovelock, J. 2000. Gaia: A New Look at Life on Earth. Oxford University Press, New York.
- Luisi, P.L. 2002. Some open questions about the origin of life. In: Fundamentals of Life. Elsevier, Paris.
- Maynard-Smith, J. and E. Szathmary. 1999. The Origins of Life: From the Birth of Life to the Origin of Language. Oxford University Press, New York.
- McMullin, B. and F.R. Varela. 1994. Rediscovering computational autopoiesis. In: Husband, P. and Harvey, I. (eds.) Proceedings of the 4th European Conference on Artificial Life, p. 38. MIT Press, Cambridge.
- Meinhardt, H. 1998. The Algorithmic Beauty of Sea Shells, 2nd enlarged edn. Springer, Berlin Heidelberg New York.
- Molter, C. and H. Bersini. 2003. Fascinating rhythms by chaotic Hopfield networks. In: Proceedings of ECAL 2003, pp. 191–198. Springer, Berlin Heidelberg New York.
- Morange, M. 2003. La vie expliquée ? 50 ans après la double hélice. Odile Jacob, Paris.
- Nagel, T. 1974. What is it Like to be a Bat? *Philosophical Review* 83. pp. 435–450. Reprinted in: Mortal Questions, pp. 165–180. Cambridge University Press, New York.
- Plasson, R., Bersini, H. and A. Commeyras. 2004. Homochirality as fixed point in prebiotic chemistry. *PNAS* 101, pp. 16733–16738.

- Poundstone, W. 1985. *The Recursive Universe*. Contemporary Books, Chicago, IL.
- Stewart, J. 1994. *The Primordial VRM System and the Evolution of Vertebrate Immunity*. Landes, Austin, TX.
- Sutton, R. and A. Barto. 1998. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge.
- Turing, A.M. 1952. The chemical basis of morphogenesis. *Philosophical Transactions the Royal Society London B* 237, pp. 37–72.
- Varela, F.R., Maturana, H.R. and R. Uribe. 1974. Autopoiesis: the organisation of living systems, its characterization and a model. *BioSystems* 5, p. 197.
- Varela, F. and A. Coutinho. 1991. Second Generation Immune Network. *Immunology Today* 12(5), pp. 159–166.
- Wolfram, S. 2002. *A New Kind of Science*. Wolfram Media, Champaign, IL.

Appendix

1 Some Astrophysical Reminders

Marc Ollivier

1.1 A Physics and Astrophysics Overview

1.1.1 Star or Planet?

Roughly speaking, we can say that the physics of stars and planets is mainly governed by their mass and thus by two effects:

1. Gravitation that tends to compress the object, thus releasing gravitational energy
2. Nuclear processes that start as the core temperature of the object increases

The mass is thus a good parameter for classifying the different astrophysical objects, the adapted mass unit being the solar mass (written M_{\odot}). As the mass decreases, three categories of objects can be distinguished:

1. if $M > 0.08 M_{\odot}$ ($\sim 80M_J$ where M_J is the Jupiter mass) the mass is sufficient and, as a consequence, the gravitational contraction in the core of the object is strong enough to start hydrogen fusion reactions. The object is then called a “star” and its radius is proportional to its mass.
2. If $0.013 M_{\odot} < M < 0.08 M_{\odot}$ ($13 M_J < M < 80 M_J$), the core temperature is not high enough for hydrogen fusion reactions, but does allow deuterium fusion reactions. The object is called a “brown dwarf” and its radius is inversely proportional to the cube root of its mass.
3. If $M < 0.013 M_{\odot}$ ($M < 13 M_J$) the temperature at the center of the object does not permit any nuclear fusion reactions. The object is called a “planet”. In this category one distinguishes giant gaseous and telluric planets. This latter is not massive enough to accrete gas. The mass limit between giant and telluric planets is about 10 terrestrial masses.

Remarks:

- Contrary to hydrogen fusion, deuterium fusion does not play a role in determining the nature of the object. The $13 M_J$ limit between a planet and a brown dwarf results from a convention (and a consensus).
- A planet is also (and first?) a body orbiting a star. Free floating brown dwarfs can be observed without necessarily being gravitationally bound to a central object.

1.1.2 Gravitation and Kepler's Laws

Kepler's three laws, even if they were formulated before Newton's theory¹, result from the universal gravitation law that postulates that two masses m_1 and m_2 separated by a distance R induce, upon one another, an attractive force \mathbf{F} parallel to the radius vector \mathbf{R} between their respective centers of mass (the so-called central force). This force is given by the relation

$$\mathbf{F} = \left(\frac{G m_1 m_2}{R^2} \right) \left(\frac{\mathbf{R}}{R} \right) \quad (1.1)$$

where G is the universal gravitation constant $= 6.67 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$.

Kepler's laws are valid for two-body systems and remain valid for multiple systems (several planets) in the planetary low mass approximation (compared to the mass of the parent star). They were derived for the solar system but can be generalized to all the planetary systems.

The first law, called the orbits law (1605) states that in the heliocentric referential, the orbit of each planet is an ellipse where the Sun is at one focus.

The second law, called the areas law, states that whatever the position in the orbit, the movement of each planet is such that the segment joining the Sun and the planet covers similar areas during the same movement duration.

The third law, called the periods law (1618) states that for each planet, the ratio between the orbit semi-major axis a cubed and the orbital period T squared is constant, with the relation

$$\frac{a^3}{T^2} = \text{const} = \frac{G(m_{\text{star}} + m_{\text{planet}})}{4\pi^2} \approx \frac{G m_{\text{star}}}{4\pi^2} \quad (1.2)$$

where m_{star} and m_{planet} are, respectively, the mass of the star and of the planet.

1.1.3 The Solar System

In the classical presentation our solar system is composed of a star (the Sun) and nine planets:² five telluric (Mercury, Venus, the Earth, Mars and Pluto) and four gaseous giants (Jupiter, Saturn, Uranus and Neptune). Table 1.1 summarizes the main characteristics of these planets.

¹ Kepler's (1571–1630) laws were empirically established beginning with detailed observations of Venus, Mars, Jupiter and Saturn's movements by Tycho Brahe (1546–1601). The first two laws were published in 1609 in *Astronomia Nova*, the third was published in 1619 in *Harmonices Mundi*. They were mathematically demonstrated by Isaac Newton (1642–1727)

² A more modern presentation would exclude Pluto from this list, because several similar objects have now been discovered near Pluto's orbit. These objects are called "transneptunians". With this new classification, Pluto would be the first and the most massive transneptunian. Observations and present solar system formation theories show that Pluto is more comparable to these objects, in terms of mass and orbit, than to Uranus and Neptune.

Table 1.1. The main characteristics of the solar system planets

Planet	Equat. Diam. (km)*	Mass (m_{\oplus})**	Distance (AU)***	Period (yr)****
Mercury	4850 (0.38)	0.0554	0.3871	0.2409
Venus	12140 (0.95)	0.815	0.7233	0.6152
Earth	12756 (1.0)	1.00	1.000	1.000
Mars	6790 (0.532)	0.1075	1.5237	1.8809
Jupiter	142600 (11.18)	317.83	5.2028	11.8623
Saturn	120200 (9.42)	95.147	9.5388	29.4577
Uranus	49000 (3.84)	14.54	19.1819	84.0139
Neptune	50200 (3.93)	17.23	30.0578	164.793
Pluto	6400 (0.52)	0.17	39.44	247.7

* in brackets, the fraction of terrestrial diameter

** m_{\oplus} = mass of the Earth = 5.976×10^{24} kg

*** AU = astronomical unit = Earth-Sun mean distance = 1.49598×10^{11} m

**** 1 yr = 365.25 terrestrial days = 8766 hours = 31 557 600 s

It is interesting to keep in mind several numbers from the previous table in order to compare exosystem characteristics, for instance:

- The mass of Jupiter is the mass unit for giant planets.
- The mass of the Earth is the mass unit for telluric planets.
- The astronomical unit expresses the distance between an (exo)planet and its central star, and is one of the main physical characteristics.
- An idea of the distance/period relation. This last point is fundamental in evaluating the observation times required to reconstruct an orbit.

1.1.4 Black Body Emission, Planck Law, Stefan–Boltzmann Law

By definition, a black body is an idealized physical body, isolated, constituted of a medium in thermodynamic equilibrium and characterized by a single equilibrium temperature. It is a perfect absorbing medium, and therefore a perfect emitting medium. The radiation field of a black body is isotropic and depends only on the temperature. The spectral distribution of the radiation intensity is given by the Planck function that gives the monochromatic brightness of the black body at a frequency ν as a function of its temperature T

$$B_{\nu} = \frac{2 h \nu^3}{c^2 e^{\left(\frac{h\nu}{kT} - 1\right)}} \text{Wm}^{-2} \text{sr}^{-1} \text{Hz}^{-1} \quad (1.3)$$

where ν is the frequency in Hz, h is Planck’s constant = 6.62620×10^{-34} J s, c is the speed of the light = 2.9979×10^8 ms⁻¹, and k is Boltzmann’s constant = 1.38×10^{-23} JK⁻¹.

One can plot the Planck function for several values of T (Fig. 1.1). This function reaches a maximum that depends on the temperature. At each temperature one can thus associate a color (wavelength of the maximum of emission) and reciprocally, after determination of the emission maximum, determine a so called black body temperature.

By integration of the Planck law over all frequencies and all directions one can derive the total power (or flux) emitted by the black body at a temperature T . This relation is known as the Stefan–Boltzmann law

$$F = \sigma T^4 W m^{-2} \tag{1.4}$$

where σ is Stefan’s constant = 5.66956×10^{-8} W m⁻² K⁻⁴.

Reciprocally, to each source emitting a flux F (measured by a bolometer, for instance) one can associate a temperature called the effective temperature T_{eff} given by the Stefan–Boltzmann law, starting from F .

A direct application of the Stefan–Boltzmann law is the calculation of the effective temperature of a planet (radius R_{pl} , mean albedo A , distance from the star D) in radiative equilibrium with a star emitting a flux S . The equality between the flux received from the star and the flux emitted by the planet can be written as a function of the preceding parameters

$$S(1 - A) \frac{\pi R_{\text{pl}}^2}{D^2} = 4 \pi R_{\text{pl}}^2 \sigma T_{\text{eff}}^4 . \tag{1.5}$$

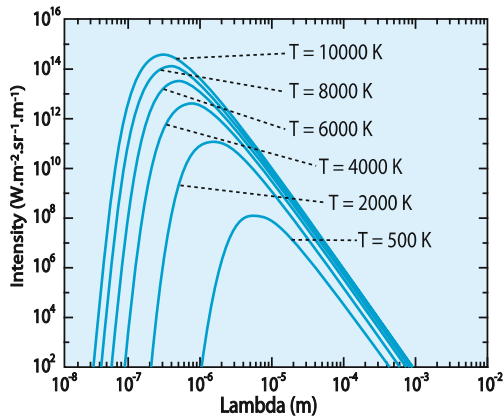


Fig. 1.1. Plot of the Planck law for several black body temperatures. The higher the black body temperature, the more the emission maximum moves towards short wavelengths

In particular, one can see that the equilibrium temperature of the planet (T_{eff}) does not depend on its size, but only on the flux from the star, the albedo and the distance from the star.

1.1.5 Hertzsprung–Russel Diagram, the Spectral Classification of Stars

The Hertzsprung–Russel diagram (called later the HR diagram) was established in 1911 by the Danish astronomer Hertzsprung and independently rediscovered in 1913 by the American astronomer Russel. This diagram plots the absolute luminosity of a star (independent of its distance to the Earth) as a function of its effective temperature or another quantity linked to the temperature (for instance, the luminosity difference of the object through two different color filters (two different bandpasses)). This diagram is shown in Fig. 1.2. This diagram is a tool to visualize, not only the stellar morphology diversity, but also stellar evolution with time. Stars have a position on the HR diagram that evolves with time, from their birth to their death.

Ninety percent of the stars that we can see are in the adult phase of their life and occupy a part of the diagram that is called the main sequence (inside the *blue ellipse* in Fig. 1.2). Our Sun, which was born five billion years ago and will die in about five billion years, is now on the main sequence. The particularity for main sequence stars is that there is a simple relation between their effective temperature, their absolute luminosity, their mass and their lifetime. As effective temperatures and colors are directly linked by the Stefan–Boltzmann law, one can see that the HR diagram is a powerful tool because it allows, starting from the object color, to arrive at other stellar characteristics. For instance,

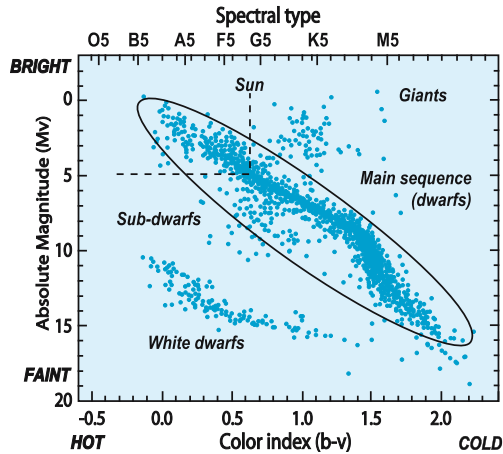


Fig. 1.2. An example of the HR diagram (the main sequence in blue)

the mean lifetime t_{life} depends on the luminosity or the mass by the following relation:

$$t_{\text{life}} \approx 10^{10} \left(\frac{m_{\text{Sun}}}{m_*} \right)^2 \approx 10^{10} \left(\frac{L_{\text{Sun}}}{L_*} \right)^{2/3} \quad (1.6)$$

where m_{Sun} , m_* , L_{Sun} , L_* are, respectively, the masses and absolute luminosities of our Sun and the star.

The HR diagram illustrates the idea of a spectral classification allowing to distinguish the stars through spectroscopic criteria. The best known and used spectral classification is the so-called Harvard classification established in 1872. This classification, based on the effective temperature of the star, defines nine categories, of which the first seven (O, B, A, F, G, K, M) include about 99% of all the stars in the sky. Table 1.2 describes the different categories of the Harvard classification. This classification is now extended to cooler objects (low-mass star to substellar objects) with the introduction of L and T types.

1. Each class is divided into 10 subclasses numbered from 0 to 9, from hotter to cooler stars (a G0 star is hotter than a G9 star).
2. The same spectral types are used to describe all stars, whatever their type. However, to differentiate between objects and because the effective temperature of a G0 giant is not the same as a G0 main sequence star, the spectral type is followed by a roman numeral describing the nature of the object. The convention is the following:

Table 1.2. Harvard spectral classification (from Léna 1996, see References in Chap. 6)

Spect. type	Star T_{eff}	Star color	Spectroscopic signatures
O	> 25000 K	Blue	Ionized helium (He II)
B	11000–25000 K	Blue-white	Neutral helium
A	7500–11000 K	White	Hydrogen
F	6000–7500 K	Yellow-white	Hydrogen (weaker than for A stars, ionized calcium
G	5000–6000 K	Yellow (~Sun)	Ionized calcium (Ca II), H, and K, CH bands
K	3500–5000 K	Orange-yellow	Metallic features
M	< 3500 K	Red	Titanium oxide (TiO) molecular bands
C (rare)	< 3500 K	Carbon stars	Molecular bands C ₂ , CN, CH, but not TiO
S (rare)	< 3500 K	Red	Zirconium oxide ZrO

- I: Supergiants
- II: Luminous giants
- III: Giants
- IV: Subgiants
- V: Dwarfs
- VI: Subdwarfs
- VII: White dwarfs

3. These categories correspond to groups easily identifiable in the HR diagram.
4. The Sun, in this classification, is a G2V star with $T_{\text{eff}} = 5777\text{K}$.

1.2 Exoplanet Detection and Characterization

1.2.1 Planet Detection by the Radial Velocity Method

As already mentioned in Chap. 6, we will not detail the various methods used to detect and analyze exoplanets. However, as almost all presently known planets were discovered by radial velocity measurements, it is interesting to understand the basics and main biases of this method.

One usually says that a planet orbits its central star. This is not exact. It is more correct to say that the whole star-planet system orbits around the center of mass of the system. If the mass of the planet were zero compared to the mass of the star, the center of the system and the center of mass of the star would be the same. In general, if m_* , m_p , O_* and O_p are, respectively, the masses and centers of the star and the planet, then the center of mass, G , of the system is given by the relation

$$\overline{O_*G} = \frac{m_p}{m_* + m_p} \overline{O_*O_p}. \quad (1.7)$$

As a consequence, and during the planet revolution around the star, the star itself moves around the point G . This movement can be detected by observing a lateral wobble of the star on the sky. This is what is derived from astrometric observations (for instance, it is the method that van de Kamp used for Barnard's star). This movement can also be detected by observing the radial wobble of the star along the line of sight. This observation can be very accurate with the use of the Doppler-Fizeau effect; the wavelength of emission of a moving object is shifted by a relative quantity v/c where v is the speed of the object, and c the speed of the light. If the object is moving closer, the wavelengths are shifted shortwards (high frequencies), if the object is moving away, the wavelengths are shifted longwards (low frequencies). This phenomenon can be illustrated by the every day example where sound waves change tonality depending on whether the vehicle is approaching or receding.

The spectroscopic observation of a star allows a precise measurement of its radial velocity (through spectral feature shifts compared to a stable spectral

reference). One thus gets the stellar radial velocity variations (Fig. 1.3), which can be written, in the case of a circular orbit³ as

$$\delta Vr = \frac{2\pi}{T} \frac{m_p}{m_*} \sin(i) a_p \approx \left(\frac{2\pi}{T}\right)^{1/3} \frac{m_p}{m_*^{2/3}} \sin(i) \quad (1.8)$$

where i is the angle at which the system is seen (if the system is seen edge-on, $I = 90^\circ$) and a_p is the semi-major axis of the planet orbit.

The preceding formula, valid for $m_p \ll m_*$, requires a few comments:

- The radial velocity method does not allow us to derive the system angle. As a consequence, the estimation of the companion mass is a lower limit. In publications, one usually finds the value of $m_p \sin(i)$. However, when the number of detected objects is high, and particularly when one looks at the mass distribution of these planets, one can use a statistical argument for the angular distribution (for instance, uniform distribution) to deconvolute the mass distribution. This method does not allow, for an individual object, the derivation of the exact mass of the companion.
- The method is most sensitive when the companion is massive and near its parent star.
- The planet can be detected only if it is observed over an important fraction of its orbit and if it induces a significant effect on the star. For objects with the same amplitude radial velocity variations, short period systems are easier to

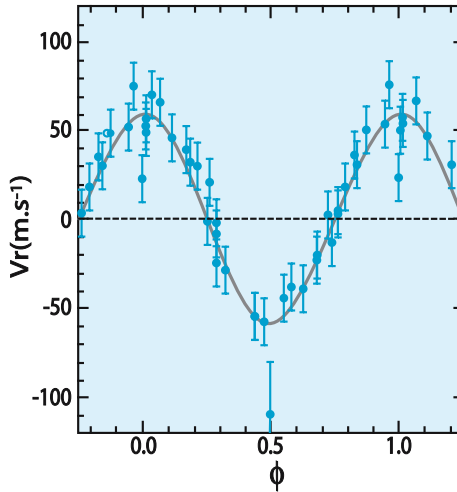


Fig. 1.3. 51 Peg radial velocity variation as a function of the orbital phase (Mayor and Queloz 1995; see References in Chap. 6)

³ The calculation in the general case of an elliptic orbit and a random orientation of the system is described in Bordé 2003 (see References in Chap. 6).

detect than long period ones. The first detected planet was 51 Peg b, which is also the first detected hot Jupiter. To date, this method has only allowed the detection of giant planets (the lightest of which has a mass of about 15 terrestrial masses, the mass of Uranus). The present instrumental limit is about $\delta Vr = 1 \text{ ms}^{-1}$. Table 1.3 gives the expected values of Vr for several example objects. This value should be associated with the revolution period of the object in order to evaluate its detectability.

- The orbital eccentricity is deduced from the radial velocity curve by the determination of its deviation from harmonicity. A perfectly circular planetary orbit induces a perfectly circular movement of the star and thus a sinusoidal variation of the radial velocity with time.
- Multiple systems are detected by a frequency analysis of the radial velocity variations, successive subtractions of periodic components and a detailed study of residuals after each subtraction.

1.2.2 Planet Detection by Astrometry

As for the radial velocity method, this method is also an indirect one. Astrometric instruments allow us to accurately measure, at a given time, the position of an object on the sky, with respect to a reference which is considered to be fixed or in uniform translation (for instance stars very far from the target) or with respect to another nearby object in the field of view. In the first case, the method is called absolute astrometry, in the other, relative astrometry. The observation of stellar position variations with time allows us to determine the exact distance to the star but also its proper motion on the sky.

In the simple case of a star single-planet system, in which the star rotates around the center of mass of the system, the proper motion of the star has an elliptic trajectory on the sky. The eccentricity mainly depends on the inclination angle i of the system, with respect to the line of sight. This is the same angle as

Table 1.3. Radial velocity variations for several example objects

Planet	m_p (m_\oplus)	a_p (AU)	a_* (r_\odot)	T (days)	δVr (ms^{-1})
Jupiter	317.83	5.20	1.07	4332.6	12.5
Saturn	95.15	9.54	0.59	10759.2	2.8
Uranus	14.54	19.18	0.18	30685.4	0.30
Neptune	17.23	30.06	0.33	60189	0.28
Earth	1	1	6.5×10^{-4}	365.25	0.09
51 Peg B	130	0.05	0.004	4.23	50.2

in the expression of the radial velocity. The angular size of the semi-major axis of this apparent orbit $\delta\theta$ can be written

$$\delta\theta = \frac{a_*}{D} = \frac{m_p}{m_*} \cdot \frac{a_p}{D} \quad (1.9)$$

where a_* , a_p are respectively the semi-major axis of the star and planet orbits, m_* and m_p are the stellar and planetary masses and D is the distance between the system and us. The term $\delta\theta$ is the astrometric instrumental accuracy required to detect the presence of the planet. Table 1.4 gives the value of $\delta\theta$ for several objects types.

The astrometric method is most effective when the companion is massive and far from its parent star (contrary to the radial velocity method, which is more sensitive to nearby companions). In this case, the amplitude of the wobble is larger. The distance between the star and the companion is deduced either from angular measurements knowing the star distance, or directly, starting from the orbital period measurement and the Kepler's third law.

In the case of multiple systems, and for two planets with similar effects, the trajectory quickly becomes complex. As an example, Fig. 1.4 shows the Sun's trajectory seen pole on, at a distance of 10 pc as a function of time. This trajectory is mainly due to the effects of Jupiter and Saturn.

1.2.3 Planet Detection by the Transit Method

When a stellar system is seen edge on (the line of sight from the Earth is in the exoplanetary system plane) one can expect to see one or several planets passing in front of the star, leading to a microeclipse that we will call a transit.

Assuming a random orientation for planetary systems the probability p that a planet of radius R_p with a revolution period T in an orbit with semi-major

Table 1.4. Amplitude of the astrometric wobble of a target star as a function of the detected object type (same mass, same distance)

Object	a_* (r_\odot)	$\delta\theta$ at 5 pc (μas)	$\delta\theta$ at 10 pc (μas)	$\delta\theta$ at 15 pc (μas)
Jupiter	1.07	1000	500	333
Saturn	0.59	550	270	180
Neptune	0.33	310	150	100
Uranus	0.18	170	80	60
Earth	$6.5 \cdot 10^{-4}$	0.6	0.3	0.2

$1 \mu\text{as} = 1 \text{ microarcsecond} \sim 5 \times 10^{-12} \text{ rad}$

axis a , transits in front of a star of radius R_* is given by the following relation (Rosenblatt⁴ 1971):

$$p = \frac{R_*}{a} . \quad (1.10)$$

The transit duration D_t is then given by the relation

$$D_t = \frac{T}{\pi} \cdot \frac{R_*}{a} . \quad (1.11)$$

This transit leads to a relative photometric extinction $\Delta F/F$ of the star given by the relation

$$\frac{\Delta F}{F} = \left(\frac{R_p}{R_*} \right)^2 . \quad (1.12)$$

Table 1.5 shows estimates of several parameters for some objects of our solar system and in other systems. For solar system objects, the transit probability must be understood as if we suppose an exoplanet with the same mass, size and distance from the star as the solar system object.

Table 1.5 clearly shows that the transit of massive objects (giant planets) should be detectable from the ground where the reachable photometric accuracy is about one percent, even a fraction of one percent in the case of very good

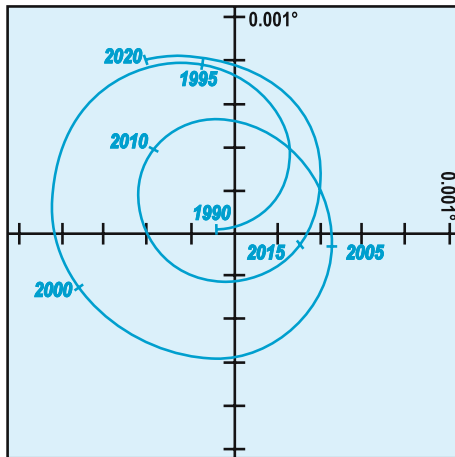


Fig. 1.4. The Sun's wobble as seen from a distance of 10 pc. One clearly distinguishes the contribution of Jupiter and other planets perturbations

⁴ In the original paper, the author overestimated the detection probability by a factor of π (corrected in this text). This paper, written more than 31 years ago by a physiologist contains a lot of ideas, particularly concerning the limb darkening effect and the chromaticity of the phenomenon. See References in Chap. 6

Table 1.5. Transit characteristics for several objects

Object	Transit probability (%)	D_t (hours)	$\Delta F/F$
Mercury ⁵	1.2	8	1.2×10^{-5}
Venus ⁶	0.64	11	7.6×10^{-5}
Earth	0.47	13	8.4×10^{-5}
Jupiter	0.09	30	10^{-2}
Saturn	0.05	41	7.5×10^{-3}
51 Peg B	9.1	3	$\sim 10^{-2}$
HD 209458 B	10.8	3	1.6×10^{-2}

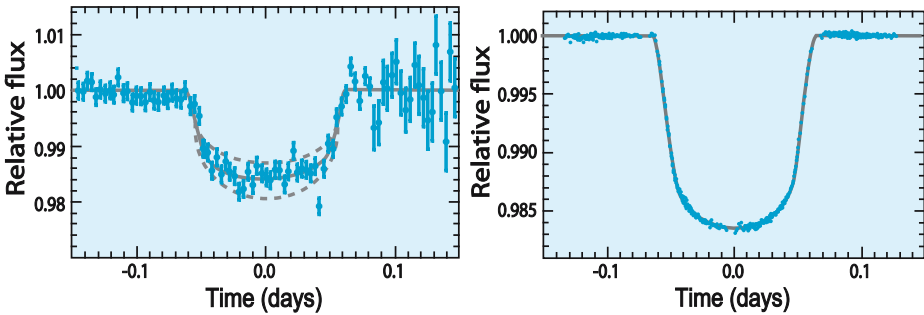


Fig. 1.5. The transit of HD 209458b from the ground (*left*) (Charbonneau et al. 2000) and (*right*) from space (Brown et al. 2001) (see References in Chap. 6)

atmospheric conditions over short duration (several hours). Low-mass object transits (telluric planets) are only detectable from space (COROT, Eddington, Kepler missions, see the main text). The absence of atmosphere in space increases the accuracy of photometric measurements. As a comparison, Fig. 1.5 shows the transit of HD 209458b observed from the ground and from space with the HST.

1.2.4 Exoplanet Direct Detection by Nulling Interferometry (Bracewell’s Interferometer)

The nulling interferometer, in its simplest version with two telescopes (also called the Bracewell interferometer after its inventor), is an instrument which combines high angular resolution and high dynamic range capabilities (Bracewell 1978; see References in Chap. 6). It is currently called an interferometric coronagraph by

⁵ Last observed transit of Venus in front of the Sun: the 8th of June 2004

⁶ Last observed transit of Mercury in front of the Sun: the 7th of May 2003

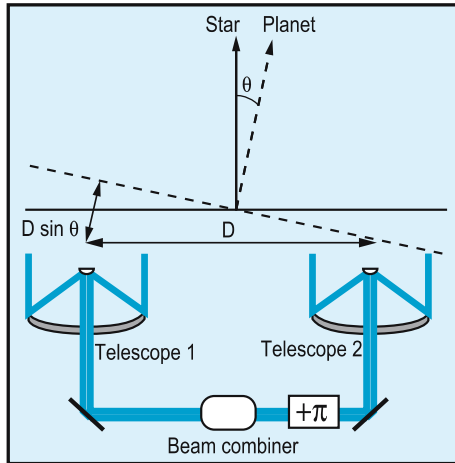


Fig. 1.6. Bracewell's interferometer (see the text for a description of the instrument)

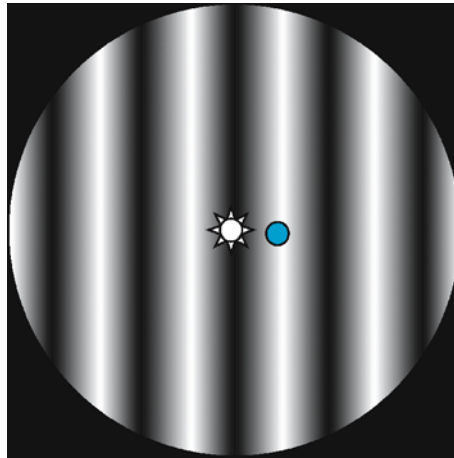


Fig. 1.7. Instrumental transmission of a Bracewell's interferometer at a wavelength of $10\mu\text{m}$ for a distance $D = 25\text{m}$ between the two telescopes. The star appears put on a dark fringe (it is cancelled) and the planet on a bright fringe. The first bright fringe corresponds to a distance of 1 AU for a star a 10 pc

analogy with Lyot's instrument. The principle of Bracewell's interferometer is shown in Fig. 1.6.

Let us consider two telescopes T_1 and T_2 , which individually, and because of diffraction, do not resolve the star-planet system. Pointing the two telescopes in the exact direction of the star and combining the beams from the two telescopes with a beam combiner, the wavefront simultaneously reaches T_1 and T_2 . When recombined the two beams will be in phase, and construc-

tive interference will be observed. If an achromatic π phase shifter is added to one of the interferometer arms (for instance, to the arm from T_2), the two beams will have opposite phase and destructive interference will be observed, thus cancelling the light coming from the star (and particularly the flux of the star). In the direction of the planet (at an angle of $\theta \sim 0.1$ arcsec from the direction of the star) one induces a delay to T_1 with respect to T_2 , equal to $D \sin(\theta)$ where D is the distance between the two telescopes. If D is adjusted (by moving the telescopes) one can get, at a mean wavelength of for instance $10\mu\text{m}$, a path difference $D \sin(\theta)$ that compensates for the π phase shift introduced in the T_2 arm. In the direction of the planet one thus gets a constructive interference. To summarize, such an instrument allows us to observe off-axis, faint objects, while the on-axis theoretical transmission is zero. The transmission of such an instrument is shown in Fig. 1.7. In its original version, Bracewell’s interferometer was rotated to modulate the signal of the planet with respect to the stellar leaks (because of the finite angular size of the star and the small size of the low transmission zone the stellar cancellation is not perfect).

1.3 List of Exoplanets as Detected the 31th of January 2005 (Schneider 2004; see References in Chap. 6)

Table 1.6. Transit characteristics for several objects

Star	Comp	$M_{\sin(i)}$ (M_{Jup})	Period (d)	Sem Maj (AU)	e
OGLE-TR-56	b	1.45	1.2	0.0225	
OGLE-TR-113	b	1.35	1.43	0.0228	
OGLE-TR-132	b	1.01	1.69	0.0306	
HD 73256	b	1.85	2.54863	0.037	0.038
GJ 436	b	0.067	2.6441	0.0278	0.12
55 Cnc	e	0.045	2.81	0.038	0.174
	b	0.84	14.65	0.11	0.02
	c	0.21?	44.28?	0.24?	0.34?
	d	4.05	5360	5.9	0.16
HD 83443	b	0.41	2.985	0.04	0.08
HD 46375	b	0.249	3.024	0.041	0.04
TrES-1	b	0.75	3.030065	0.0393	0.0
HD 179949	b	0.84	3.093	0.045	0.05
HD 187123	b	0.52	3.097	0.042	0.03
OGLE-TR-10	B	0.57	3.101386	0.04162	

Table 1.6. (continued)

Star	Comp	$M_{\sin(i)}$ (M_{Jup})	Period (d)	Sem Maj (AU)	e
Tau Boo	b	3.87	3.3128	0.0462	0.018
HD 330075	b	0.76	3.369	0.043	0
HD 88133	b	0.29	3.415	0.046	0.11
BD-10 3166	b	0.48	3.487	0.046	0
HD 75289	b	0.42	3.51	0.046	0.054
HD 209458	b	0.69	3.5247542	0.045	0
HD 76700	b	0.197	3.971	0.049	0
OGLE-TR-111	b	0.53	4.02	0.047	
51 Peg	b	0.468	4.23077	0.052	0
Ups And	b	0.69	4.617	0.059	0.012
	c	1.19	241.5	0.829	0.28
	d	3.75	1284	2.53	0.27
HD 49674	b	0.12	4.948	0.0568	0
HD 68988	b	1.9	6.276	0.071	0.14
HD 168746	b	0.23	6.403	0.065	0.081
HD 217107	b	1.28	7.11	0.07	0.14
HD 162020	b	13.75	8.428198	0.072	0.277
HD 160691	d	0.042	9.55	0.09	0
	b	1.7	638	1.5	0.31
	c	3.1	2986	4.17	0.8?
HD 130322	b	1.08	10.724	0.088	0.048
HD 108147	b	0.41	10.901	0.104	0.498
HD 38529	b	0.78	14.309	0.129	0.29
	c	12.7	2174.3	3.68	0.36
Gl 86	b	4	15.78	0.11	0.046
HD 99492	b	0.112	17.038	0.119	0.05
HD 195019	b	3.43	18.3	0.14	0.05
HD 6434	b	0.48	22.09	0.15	0.3
HD 192263	b	0.72	24.348	0.15	0
Gliese 876	c	0.56	30.1	0.13	0.27
	b	1.98	61.02	0.21	0.12
HD 102117	b	0.18	20.8	0.15	0.08
HD 11964	c	0.11	37.82	0.23	0.15
	b	0.7	1940	3.17	0.3
Rho Crb	b	1.04	39.845	0.22	0.04
HD 74156	b	1.86	51.643	0.294	0.636

Table 1.6. (continued)

Star	Comp	$M_{\sin(i)}$ (M_{Jup})	Period (d)	Sem Maj (AU)	e
	c	> 6.17	2025	3.4	0.583
HD 117618	b	0.22	52	0.28	0.25
HD 37605	b	2.85	55.2	0.26	0.736
HD 168443	b	7.7	58.116	0.29	0.529
	c	16.9	1739.5	2.85	0.228
HD 3651	b	0.2	62.23	0.284	0.63
HD 121504	b	0.89	64.6	0.32	0.13
HD 178911B	b	6.292	71.487	0.32	0.1243
HD 16141	b	0.23	75.56	0.35	0.28
HD 114762	b	11	84.03	0.3	0.334
HD 80606	b	3.41	111.78	0.439	0.927
70 Vir	b	7.44	116.689	0.4	0.4
HD 216770	b	0.65	118.45	0.46	0.37
HD 52265	b	1.13	118.96	0.49	0.29
HD 34445	b	0.58	126	0.51	0.4
HD 208487	b	0.41	129.1	0.52	0.35
GJ 3021	b	3.21	133.82	0.49	0.505
HD 37124	b	0.75	152.4	0.54	0.1
	c	1.2	1495	2.5	0.69
HD 219449	b	2.9	182	0.3	
HD 73526	b	3	190.5	0.66	0.34
HD 104985	b	6.3	198.2	0.78	0.03
HD 82943	b	0.88	221.6	0.73	0.54
	c	1.63	444.6	1.16	0.41
HD 169830	b	2.88	225.62	0.81	0.31
	c	4.04	2102	3.6	0.33
HD 8574	b	2.23	228.5	0.76	0.4
HD 202206	c	2.44	1383.4	2.55	0.267
HD 89744	b	7.99	256.6	0.89	0.67
HD 134987	b	1.58	260	0.78	0.25
HD 40979	b	3.32	267.2	0.811	0.25
HD 12661	b	2.3	263.6	0.83	0.096
	c	1.57	1444.5	2.56	< 0.1
HD 150706	b	1	264.9	0.82	0.38
HD 59686	b	6.5	303	0.8	
HR 810	b	2.26	320.1	0.925	0.161

Table 1.6. (continued)

Star	Comp	$M_{\sin(i)}$ (M_{Jup})	Period (d)	Sem Maj (AU)	e
HD 142	b	1.36	338	0.98	0.37
HD 92788	b	3.8	340	0.94	0.36
HD 28185	b	5.6	385	1	0.06
HD 196885	b	1.84	386	1.12	0.3
HD 142415	b	1.62	386.3	1.05	0.5
HD 177830	b	1.28	391	1	0.43
HD 154857	b	1.8	398	1.11	0.51
HD 108874	b	1.65	401	1.07	0.2
HD 4203	b	1.65	400.994	1.09	0.46
HD 128311	b	2.58	420	1.02	0.3
	c	3.24	919	1.74	0.29
HD 27442	b	1.28	423.841	1.18	0.07
HD 210277	b	1.28	437	1.097	0.45
HD 19994	b	2	454	1.3	0.2
HD 188015	b	1.64	471	1.22	0.26
HD 20367	b	1.07	500	1.25	0.23
HD 114783	b	0.9	501	1.2	0.1
HD 147513	b	1	540.4	1.26	0.52
HIP 75458	b	8.64	550.651	1.34	0.71
HD 222582	b	5.11	572	1.35	0.76
HD 65216	b	1.21	613.1	1.37	0.41
HD 183263	b	3.77	631.45	1.52	0.37
HD 141937	b	9.7	653.22	1.52	0.41
HD 41004 A	b	2.3	655	1.31	0.39
HD 47536	b	4.96–9.67	712.13	1.61–2.25	0.2
HD 23079	b	2.61	738.459	1.65	0.1
16 Cyg B	b	1.69	798.938	1.67	0.67
HD 4208	b	0.80	812.197	1.67	0.05
HD 114386	b	0.99	872	1.62	0.28
HD 45350	b	1.03	890	1.86	0.78
Gamma Cephei	b	1.59	902.96	2.03	0.2
HD 213240	b	4.5	951	2.03	0.45
HD 10647	b	0.91	1040	2.1	0.18
HD 10697	b	6.12	1077. 906	2.13	0.11
47 Uma	b	2.41	1095	2.1	0.096
	c	0.76	2594	3.73	< 0.1

Table 1.6. (continued)

Star	Comp	$M_{\sin(i)}$ (M_{Jup})	Period (d)	Sem Maj (AU)	e
HD 190228	b	4.99	1127	2.31	0.43
HD 114729	b	0.82	1131.478	2.08	0.31
HD 111232	b	6.8	1143	1.97	0.20
HD 2039	b	4.85	1192.582	2.19	0.68
HD 136118	b	11.9	1209.6	2.335	0.366
HD 50554	b	4.9	1279	2.38	0.42
HD 196050	b	3	1289	2.5	0.28
HD 216437	b	2.1	1294	2.7	0.34
HD 216435	b	1.49	1442.919	2.7	0.34
HD 106252	b	6.81	1500	2.61	0.54
HD 23596	b	7.19	1558	2.72	0.314
14 Her	b	4.74	1796.4	2.8	0.338
OGLE-235/ MOA-53	b	1.5–2.5	?	2.8–3	?
HD 39091	b	10.35	2063.818	3.29	0.62
HD 72659	b	2.55	2185	3.24	0.18
HD 70642	b	2	2231	3.3	0.1
HD 33636	b	9.28	2447.292	3.56	0.53
Eps Eri	b	0.86	2502.1	3.3	0.608
	c	0.1?	102270?	40?	0.3?
HD 117207	b	2.1	2617	3.8	0.16
HD 30177	b	9.17	2819.654	3.86	0.3
G1 777A	b	1.33	2902	4.8	0.48
HD 89307	b	2.73	3090	4.15	0.32
2M1207	b	~5	> 894250	> 55	?

2 Useful Astrobiological Data

2.1 Physical and Chemical Data

1. International System Units

	Name	Symbol	
Basic units			
Length	Metre	m	
Mass	Kilogram	kg	
Time	Second	s	
Electric current	Ampere	A	
Thermodynamic temperature	Kelvin	K	
Amount of substance	Mole	mol	
Luminous intensity	Candela	cd	
Additional units			
Plane angle	Radian	rad	
Solid angle	Steradian	sr	
Derived units			
Energy	Joule	J	$\text{m}^2 \text{kg s}^{-2} \equiv \text{N m}$
Force	Newton	N	$\text{m kg s}^{-2} \equiv \text{J m}^{-1}$
Pressure	Pascal	Pa	$\text{m}^{-1} \text{kg s}^{-2} \equiv \text{N m}^{-2}$
Electrical charge	Coulomb	C	s A
Magnetic flux	Weber	Wb	$\text{m}^2 \text{kg s}^{-2} \text{A}^{-1}$
Magnetic flux density	Tesla	T	$\text{kg s}^{-2} \text{A}^{-1} \equiv \text{V s m}^{-2}$
Magnetic field strength			A m^{-1}
Inductance	Henry	H	$\text{m}^2 \text{kg s}^{-2} \text{A}^{-2}$
Power	Watt	W	$\text{m}^2 \text{kg s}^{-2} \equiv \text{J s}^{-1}$
Electric potential difference	Volt	V	$\text{m}^2 \text{kg s}^{-3} \text{A} \equiv \text{J C}^{-1}$
Frequency	Hertz	Hz	s^{-1}

2. Other Units

	Name	Symbol	
Length	Fermi	fm	10^{-15} m
	Angstrom	Å	10^{-10} m
	Micron	μm	10^{-6} m
Time	Minute	min, mn	60 s
	Hour	h	3600 s
	Day	d	86 400 s
	Year	a, y	3.156×10^7 s
Volume	Litre	l	10^{-3} m ³ ≡ 1 dm ³
Mass	Ton	t	10 ³ kg
Temperature	Degree Celsius	°C	Temperature comparison: $0^\circ\text{C} = 273.15\text{K} = 32^\circ\text{F}$ $100^\circ\text{C} = 373.15\text{K} = 212^\circ\text{F}$ Temperature scale: $\text{K} = ^\circ\text{C} + 273.15$ $\text{K} = ^\circ\text{C} + 1.8^\circ\text{F}$
Force	Dyne (cgs system)	dyn	10^{-5} N
Energy	Erg (cgs system)	erg	10^{-7} J
	Calorie	cal	4.184 J
	Electron-volt	eV	1.6×10^{-19} J
	Kilowatt-hour		3600×10^3 J = 8.6042×10^5 cal
Pressure	Bar	bar	10^5 Pa
	Atmosphere	atm	1.01325×10^5 Pa
	mercury mm	torr	1.3332×10^2 Pa
Magnetic flux density	Gauss	gauss	10^{-4} T
Concentration	Molarity	M	mol l ⁻¹
Molar mass			g/mol
Angle	Degree	°	$1^\circ = \pi/180$
			$= 1.74 \times 10^{-2}$ rad
	Minute	'	$1' = \pi/10\ 800$
			$= 2.91 \times 10^{-4}$ rad
	Second	"	$1'' = \pi/648\ 000$
			$= 4.85 \times 10^{-6}$ rad
			1 Radian = 57.296° $= 3.44 \times 10^3$ ' $= 2.06 \times 10^5$ ''

3. International System Prefixes

10^{-1}	deci	d	10^1	deca	da
10^{-2}	centi	c	10^2	hecto	h
10^{-3}	milli	m	10^3	kilo	k
10^{-6}	micro	μ	10^6	mega	M
10^{-9}	nano	n	10^9	giga	G
10^{-12}	pico	p	10^{12}	tera	T
10^{-15}	femto	f	10^{15}	peta	P
10^{-18}	atto	a	10^{18}	exa	E
10^{-21}	zepto	z	10^{21}	zetta	Z
10^{-24}	yocto	y	10^{24}	yotta	Y

4. Fundamental Physical Constants

Newtonian constant of gravitation	G	$6.67259(85) \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$
Acceleration of free fall on Earth	g	9.80665 ms^{-2}
Speed of light in vacuum	c	$2.99792458 \times 10^8 \text{ m s}^{-1}$
Planck constant	h	$6.6260755(40) \times 10^{-34} \text{ J s}$
Planck mass	$(hc/2\pi G)^{1/2}$	$2.17671(14) \times 10^{-8} \text{ kg}$
Planck length	$(hG/2\pi c^3)^{1/2}$	$1.61605(10) \times 10^{-35} \text{ m}$
Planck time	$(hG/2\pi c^5)^{1/2}$	$5.39056(34) \times 10^{-44} \text{ s}$
Boltzmann constant	k	$1.380658(12) \times 10^{-23} \text{ J K}^{-1}$
Black body constant	a	$7.564 \times 10^{-16} \text{ J m}^{-3} \text{ K}^{-4}$
Stefan Boltzmann constant	σ	$5.67051(19) \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$
Perfect gas constant	R	$8.314510(70) \text{ J K}^{-1} \text{ mol}^{-1}$
Elementary charge	e	$1.602176462(63) \times 10^{-19} \text{ C}$
Electron rest mass	m_e	$9.1093897(54) \times 10^{-31} \text{ kg}$
Proton rest mass	m_p	$1.6726231(10) \times 10^{-27} \text{ kg}$
Neutron rest mass	m_n	$1.6749286(10) \times 10^{-27} \text{ kg}$
Proton/electron mass ratio		1836.152701(37)
Hydrogen atomic mass		$1.6735344 \times 10^{-27} \text{ kg}$
Electron rest mass energy equivalent (eV)	$m_e c^2$	$0.511 \times 10^6 \text{ eV}$

(continued)

Atomic mass unit	u	$1.6605402(10) \times 10^{-27} \text{ kg} \equiv$ $1/12 \text{ of } ^{12}\text{C mass}$
Avogadro constant	N	$6.0221367(36) \times 10^{23} \text{ mol}^{-1}$
Vacuum permittivity constant	$1/(\mu_0 c^2)$	$8.854187817 \times$ $10^{-12} \text{ m}^{-3} \text{ kg}^{-1} \text{ s}^4 \text{ A}^2$
Vacuum permeability constant	μ_0	$4\pi 10^{-7} \text{ m kg s}^{-2} \text{ A}^{-2}$
Ice melting temperature ($P = 1 \text{ atm}$)	0°C	273.150 K
Water triple point (H_2O)		$T = 273.160 \text{ K},$ $P = 611.73 \text{ Pa (6.11 mbar)}$
Bohr radius	a_0	$0.529177249(24) \times 10^{-10} \text{ m}$
Electron volt-energy relationship		$1 \text{ eV} = 1.60217733(49) \times 10^{-19} \text{ J}$
Electron volt-wavelength relationship		$1 \text{ eV} = 12\,398.428 \times 10^{-10} \text{ m}$
Electron volt-frequency relationship		$1 \text{ eV} = 2.41798836(72) \times 10^{14} \text{ s}^{-1}$
Electron volt-temperature relationship		$1 \text{ eV} = 11\,604.45 \text{ K}$

5. Other Physical, Chemical and Astronomical Symbols and Abbreviations

c	Mass concentration Light speed
E_a	Activation energy
k	Reaction rate
R	“Rectus” configuration (chirality)
S	“Sinister” configuration (chirality)
L,D	Old nomenclature still used to define the absolute configuration of oses and amino acids
δ	Chemical shift (RMN) Fractional charge
λ	Wavelength
μ	Dipolar moment Micron Reduced mass
ν	Frequency
[]	Concentration (mol l ⁻¹)
\rightleftharpoons	Chemical equilibrium
\leftrightarrow	Mesomery
UV	Ultraviolet
IR	Infrared
NMR	Nuclear Magnetic Resonance
a	Ellipse semi great axis
e	Ellipse eccentricity
i	Inclination
PSR	Pulsar
QSO	Quasar
γ	Photon
ν	Neutrino
\odot	Sun
♁	Mercury
♀	Venus
\oplus	Earth
♂	Mars
♃	Jupiter
♄	Saturn
♅	Uranus
♆	Neptune
♇	Pluto

References: *Allen's: Astrophysical quantities*, Arthur N. Cox Editor, (2000); *Introduction aux éphémérides astronomiques*, EDP Sciences, (1998); *Handbook of Chemistry and Physics*, D.R. Lide, editor-in-Chief 82nd Edition (2001–2002) CRC Press.

										18 VIIIA					
										Helium 2 4.002602 1s ² He					
										13 IIIA	14 IVA	15 VA	16 VIA	17 VIIA	18 VIIIA
										Boron 5 10.811 1s ² 2s ² 2p ¹ B	Carbon 6 12.011 1s ² 2s ² 2p ² C	Nitrogen 7 14.00674 1s ² 2s ² 2p ³ N	Oxygen 8 15.9994 1s ² 2s ² 2p ⁴ O	Fluorine 9 18.9984032 1s ² 2s ² 2p ⁵ F	Neon 10 20.1797 1s ² 2s ² 2p ⁶ Ne
										Aluminium 13 26.981539 [Ne] 3s ² 3p ¹ Al	Silicium 14 28.0855 [Ne] 3s ² 3p ² Si	Phosphorus 15 30.973762 [Ne] 3s ² 3p ³ P	Sulfur 16 32.066 [Ne] 3s ² 3p ⁴ S	Chlorine 17 35.453 [Ne] 3s ² 3p ⁵ Cl	Argon 18 39.948 [Ne] 3s ² 3p ⁶ Ar
10	11 IB		12 IIB												
Nickel 28 58.6934 [Ar] 3d ⁸ 4s ² Ni	Copper 29 63.546 [Ar] 3d ¹⁰ 4s ¹ Cu	Zinc 30 65.39 [Ar] 3d ¹⁰ 4s ² Zn	Gallium 31 69.723 [Ar] 3d ¹⁰ 4s ² 4p ¹ Ga	Germanium 32 72.61 [Ar] 3d ¹⁰ 4s ² 4p ² Ge	Arsenic 33 74.92159 [Ar] 3d ¹⁰ 4s ² 4p ³ As	Selenium 34 78.96 [Ar] 3d ¹⁰ 4s ² 4p ⁴ Se	Bromine 35 79.904 [Ar] 3d ¹⁰ 4s ² 4p ⁵ Br	Krypton 36 83.80 [Ar] 3d ¹⁰ 4s ² 4p ⁶ Kr	Xenon 54 131.29 [Kr] 4d ¹⁰ 5s ² 5p ⁶ Xe						
Palladium 46 106.42 [Kr] 4d ¹⁰ Pd	Silver 47 107.8682 [Kr] 4d ¹⁰ 5s ¹ Ag	Cadmium 48 112.411 [Kr] 4d ¹⁰ 5s ² Cd	Indium 49 114.82 [Kr] 4d ¹⁰ 5s ² 5p ¹ In	Tin 50 118.71 [Kr] 4d ¹⁰ 5s ² 5p ² Sn	Antimony 51 121.757 [Kr] 4d ¹⁰ 5s ² 5p ³ Sb	Tellurium 52 127.60 [Kr] 4d ¹⁰ 5s ² 5p ⁴ Te	Iodine 53 126.90447 [Kr] 4d ¹⁰ 5s ² 5p ⁵ I	Xenon 54 131.29 [Kr] 4d ¹⁰ 5s ² 5p ⁶ Xe	Radon 86 222.0176 [Xe] 4f ¹⁴ 5d ¹⁰ 6s ² 6p ⁶ Rn						
Platinum 78 195.08 [Xe] 4f ¹⁴ 5d ⁹ 6s ¹ Pt	Gold 79 196.96654 [Xe] 4f ¹⁴ 5d ¹⁰ 6s ¹ Au	Mercury 80 200.59 [Xe] 4f ¹⁴ 5d ¹⁰ 6s ² Hg	Thallium 81 204.3833 [Xe] 4f ¹⁴ 5d ¹⁰ 6s ² 6p ¹ Tl	Lead 82 207.2 [Xe] 4f ¹⁴ 5d ¹⁰ 6s ² 6p ² Pb	Bismuth 83 208.98037 [Xe] 4f ¹⁴ 5d ¹⁰ 6s ² 6p ³ Bi	Polonium 84 209 [Xe] 4f ¹⁴ 5d ¹⁰ 6s ² 6p ⁴ Po	Astatine 85 209 [Xe] 4f ¹⁴ 5d ¹⁰ 6s ² 6p ⁵ At	Radon 86 222.0176 [Xe] 4f ¹⁴ 5d ¹⁰ 6s ² 6p ⁶ Rn	Radon 86 222.0176 [Xe] 4f ¹⁴ 5d ¹⁰ 6s ² 6p ⁶ Rn						
Europium 63 151.965 [Xe] 4f ⁷ 6s ² Eu	Gadolinium 64 157.25 [Xe] 4f ⁷ 5d ¹ 6s ² Gd	Terbium 65 158.92534 [Xe] 4f ⁹ 6s ² Tb	Dysprosium 66 162.50 [Xe] 4f ¹⁰ 6s ² Dy	Holmium 67 164.93032 [Xe] 4f ¹¹ 6s ² Ho	Erbium 68 167.26 [Xe] 4f ¹² 6s ² Er	Thulium 69 168.93421 [Xe] 4f ¹³ 6s ² Tm	Ytterbium 70 173.04 [Xe] 4f ¹⁴ 6s ² Yb	Lutetium 71 174.967 [Xe] 4f ¹⁴ 5d ¹ 6s ² Lu	Lutetium 71 174.967 [Xe] 4f ¹⁴ 5d ¹ 6s ² Lu						
Americium 95 243.0614 [Rn] 5f ⁷ 7s ² Am	Curium 96 247.0703 [Rn] 5f ⁸ 7s ² Cm	Berkelium 97 247.0703 [Rn] 5f ⁹ 7s ² Bk	Californium 98 251.0796 [Rn] 5f ¹⁰ 7s ² Cf	Einsteinium 99 252.08329 [Rn] 5f ¹¹ 7s ² Es	Fermium 100 257.0951 [Rn] 5f ¹² 7s ² Fm	Mendelevium 101 258.0986 [Rn] 5f ¹³ 7s ² Md	Nobelium 102 259.1009 [Rn] 5f ¹⁴ 7s ² No	Lawrencium 103 260.1053 [Rn] 5f ¹⁴ 6d ¹ 7s ² Lr	Lawrencium 103 260.1053 [Rn] 5f ¹⁴ 6d ¹ 7s ² Lr						
NON-METALS															
□	Hydrogen		■	Halogens											
□	Other Non-Metals		■	Noble Gases											
					STANDARD CONDITIONS (P = 100 KPa; T = 298 K)		Gaz		Liquid						
							Solid		Synthetic						

2.2 Astrophysical Data

1. Units and General Data

Distance units

Astronomical unit (AU)	$1.4959787066 \times 10^{11}$ m
1 parsec (pc)	206 264.8 au
	3.0856776×10^{16} m
	3.2615638 light-years
1 light-year	9.460730×10^{15} m
	6.324×10^4 AU

Time units

1 sidereal day	23 h 56 mn 04.098 s (average time)
1 tropic year	365.2422 average solar days
1 sidereal year	365.2564 average solar days

Known universe

Number of nucleons	$\sim 10^{80}$
Universe radius	$\sim 10^{26}$ m
Number of galaxies	$\sim 10^{11}$
Nebula recession speed	$\sim 3 \times 10^{-18}$ s ⁻¹ (~ 100 (km/s)/Mpc)

Our galaxy

Number of stars	$\sim 1.6 \times 10^{11}$
Diameter	$\sim 10^{21}$ m
Mass	$\sim 8 \times 10^{41}$ kg

Sun

Radius	6.95508×10^8 m
Mass	1.9891×10^{30} kg
Sun luminosity	$3.845(8) \times 10^{26}$ W
Core temperature	1.557×10^7 K
Core pressure	2.334×10^{16} Pa

(continued)

Photosphere temperature	5780 K
Corona temperature	2 to 3×10^6 K
Average mass density	$1.41 \times 10^3 \text{ kg m}^{-3}$
Earth/Moon	
Earth's mass	5.9742×10^{24} kg
Moon's mass	7.34×10^{22} kg
Earth's average radius	6371.23×10^3 m
Earth's equatorial radius	6378.136×10^3 m
Earth–Moon distance (semi-major axis)	3.844×10^8 m
Earth–Sun distance (semi-major axis)	1.496×10^{11} m

2. Compared Planetology

Object		Semi major axis of the orbit ¹		Revolution period (d: day, yr: year, terrestrial)	Eccentricity (e)
		AU	Million km		
Mercury	♿	0.3871	57.9092	87.969 d	0.2056
Venus	♀	0.7233	108.2089	224.701 d	0.0068
Earth	♁	1.0000	149.65979	365.2564 d	0.0167
Moon		2.570×10^{-3}	0.3844	27.3217 d	0.0549
Mars	♂	1.5234	227.9364	686.98 d	0.0934
Jupiter	♃	5.2034	778.4120	11.862 yr	0.0484
Europe		4.4851×10^{-3}	0.671	3.551181 d	0.009
Saturn	♄	9.5371	1426.7254	29.457 yr	0.0542
Titan		8.167×10^{-3}	1.2218	15.945 d	0.0291
Uranus	♅	19.1913	2870.9722	84.019 yr	0.0472
Neptune	♆	30.0690	4498.2529	164.767 yr	0.0086
Pluto	♇	39.4817	5906.3762	247.688 yr	0.2488

¹ around the Sun or, in the case of satellites, around their planet

Object		Equatorial radius km	Mass $\oplus = 1$	Average density g cm^{-3}	Surface gravity ¹ m s^{-2}	Surface escape velocity ¹ km s^{-1}
Sun	☉	695 508	332 946.0	1.41	274.0	617.5
Mercury	☿	2439.7	0.0553	5.43	3.72	4.25
Venus	♀	6051.8	0.8150	5.24	8.87	10.36
Earth	♁	6378.136	1.000	5.515	9.81	11.18
Moon		1737.4	0.012300	3.34	1.62	2.38
Mars	♂	3397	0.1074	3.94	3.71	5.02
Jupiter	♃	71 492	317.82	1.33	23.12	59.54
Europe		1565	0.008026	3.04	1.8	2.0
Saturn	♄	60 268	95.161	0.70	8.96	35.49
Titan		2575	0.02226	1.90	0.14	2.7
Uranus	♅	25 559	14.371	1.30	8.69	21.29
Neptune	♆	24 764	17.147	1.76	11.0	23.71
Pluto	♇	1737.4	0.0022	1.1	0.81	1.27

¹ or at $P = 1$ bar for gaseous planets

Object		Mean obliquity (degree)	Visual geometric Albedo	Sidereal rotation period (d; day, h; hour)
Sun	☉		–	25.7 d
Mercury	☿	0.0	0.11	58.6462 d
Venus	♀	177.3	0.65	243.0185 d ¹
Earth	♁	23.45	0.37	23.9345 h
Moon		6.683	0.12	27.3217 d
Mars	♂	25.19	0.15	24.6230 h
Jupiter	♃	3.12	0.52	9.9249 h
Europe		–	0.67	2.5512 d
Saturn	♄	26.73	0.47	10.6562 h
Titan		–	0.22	15.9454 h
Uranus	♅	97.86	0.51	17.2400 h ¹
Neptune	♆	29.58	0.51	16.1100 h
Pluto	♇	119.61	0.3	153.29 h ¹

¹ retrograde rotation

3. Electromagnetic Spectrum

Wavelength	Frequency	Wave number	Equivalent energy	Equivalent temperature	Spectral Region
1000 km	300 Hz				
100 km	3 kHz			ULF	
1 km	300 kHz				
10 m	30 MHz				Radio
1 m	300 MHz			VHF	
10 cm	3 GHz			UHF	
					centimetric
1 cm	30 GHz	1 cm^{-1}			Microwave
1 mm	300 GHz	10 cm^{-1}	1.2 meV	3 K	millimetric
100 μm	3 THz	100 cm^{-1}	0.012 eV	30 K	submillimetric
				FIR	
40 μm					
10 μm	30 THz	1000 cm^{-1}	0.12 eV	300 K	Infrared
5 μm					

(continued)

Wavelength	Frequency	Wave number	Equivalent energy	Equivalent temperature	Spectral Region
1 μm	3×10^{14} Hz	$10\,000\text{ cm}^{-1}$	1.24 eV	3000 K	NIR
0.7 μm					
0.4 μm					Visible
0.1 μm	3×10^{15} Hz		12.4 eV	3×10^4 K	Ultraviolet
10 nm	3×10^{16} Hz		124 eV	3×10^5 K	
1 nm	3×10^{17} Hz		1.24 keV	3×10^6 K	
1 \AA	3×10^{18} Hz		12.4 keV	3×10^7 K	X-rays
10^{-11} m	3×10^{19} Hz		124 keV		High Energies
10^{-12} m	3×10^{20} Hz		1.24 MeV		Gamma-rays

Remarks:

- Wave numbers, energies and temperatures can be calculated for all existing wavelengths (here, they have been intentionally limited to the domains where they are generally used).
- Spectrum domain limits are no more than indicative and correspond to astronomical use.
- ULF: Ultra Low Frequency, VHF: Very High Frequency, UHF: Ultra High Frequency, FIR: Far Infrared, MIR: Mid Infrared, NIR: Near Infrared, UV: Ultraviolet
- Formula: $E = h\nu = hc/\lambda = \text{eV} = \text{kT}$

2.3 Geological Data

1. General Information

Whole Earth

Earth's Age	4.55 Ga
Age of the oldest known terrestrial material (zircon from Jack Hills; Australia)	4.404 Ga
Age of the oldest known terrestrial rock (Acasta gneiss; Canada)	4.03 Ga
Equatorial radius	6378.136 km
Polar radius	6356.753 km
Ellipticity	0.00335281
Total surface	$510 \times 10^6 \text{ km}^2$
Ocean surface	$357 \times 10^6 \text{ km}^2$
Continent surface	$153 \times 10^6 \text{ km}^2$
Mass	$5.9742 \times 10^{24} \text{ kg}$
Mean density	5.515
Deeper oceanic trench (Marianas)	11 034 m
Higher mountain (Everest)	8863 m
Present day oceanic ridge length	~80 000 km

Solid inner core

Depth	5155 km
Thickness	~1223 km
Average density	12
Volume	$8.45 \times 10^{10} \text{ km}^3$
Mass	$1.18 \times 10^{23} \text{ kg}$

Liquid outer core

Depth	2891 km
Thickness	~2264 km
Average density	10
Volume	$1.65 \times 10^{11} \text{ km}^3$
Mass	$1.65 \times 10^{24} \text{ kg}$

Lower mantle (Mesosphere)

Depth	660 km
Thickness	~2231 km
Average density	5
Volume	$5.87 \times 10^{11} \text{ km}^3$
Mass	$2.93 \times 10^{24} \text{ kg}$

(continued)

Upper mantle	
Depth	30 km
Thickness	~630 km
Average density	3.2
Volume	$3.08 \times 10^{11} \text{ km}^3$
Mass	$0.986 \times 10^{24} \text{ kg}$
Oceanic crust	
Average thickness	5 km
Average density	3.1
Age of the oldest known oceanic crust	0.165 Ga
Volume	$2.0 \times 10^9 \text{ km}^3$
Mass	$6.2 \times 10^{21} \text{ kg}$
Continental crust	
Average thickness	30 km
Average density	2.75
Age of the oldest known continental crust	4.01 Ga
Volume	$7 \times 10^9 \text{ km}^3$
Mass	$1.9 \times 10^{22} \text{ kg}$
Average surface magnetic flux density	$5 \times 10^{-5} \text{ T}$
Heat flux	$42 \times 10^{12} \text{ W}$
Range of lithospheric plate motion rate	1 to 10 cm yr ⁻¹

2. Decay Constants (λ)

Parent nuclide	Daughter nuclide	Decay constant
^{228}Th	^{224}Ra	$3.63 \times 10^{-1} \text{ yr}^{-1}$
^{210}Pb	^{210}Bi	$3.11 \times 10^{-2} \text{ yr}^{-1}$
^{32}Si	^{32}P	$2.1 \times 10^{-3} \text{ yr}^{-1}$
^{226}Ra	^{222}Rn	$4.33 \times 10^{-4} \text{ yr}^{-1}$
^{14}C	^{14}N	$1.245 \times 10^{-4} \text{ yr}^{-1}$

(continued)

Parent nuclide	Daughter nuclide	Decay constant
^{231}Pa	^{227}Ac	$2.11 \times 10^{-5} \text{ yr}^{-1}$
^{230}Th	^{226}Ra	$9.21 \times 10^{-6} \text{ yr}^{-1}$
^{59}Ni	^{59}Co	$9.12 \times 10^{-6} \text{ yr}^{-1}$
^{41}Ca	^{41}K	$6.93 \times 10^{-6} \text{ yr}^{-1}$
^{81}Kr	^{81}Br	$3.03 \times 10^{-6} \text{ yr}^{-1}$
^{234}U	^{230}Th	$2.83 \times 10^{-6} \text{ yr}^{-1}$
^{36}Cl	^{36}Ar	$2.30 \times 10^{-6} \text{ yr}^{-1}$
^{26}Al	^{26}Mg	$9.80 \times 10^{-7} \text{ yr}^{-1}$
^{107}Pd	^{107}Ag	$6.5 \times 10^{-7} \text{ yr}^{-1}$
^{60}Fe	^{60}Ni	$4.62 \times 10^{-7} \text{ yr}^{-1}$
^{10}Be	^{10}B	$4.59 \times 10^{-7} \text{ yr}^{-1}$
^{182}Hf	^{182}W	$7.7 \times 10^{-8} \text{ yr}^{-1}$
^{129}I	^{129}Xe	$4.3 \times 10^{-8} \text{ yr}^{-1}$
^{92}Nb	^{92}Zr	$1.93 \times 10^{-8} \text{ yr}^{-1}$
^{53}Mn	^{53}Cr	$1.87 \times 10^{-8} \text{ yr}^{-1}$
^{244}Pu	$^{131-136}\text{Xe}$	$8.66 \times 10^{-9} \text{ yr}^{-1}$
^{235}U	^{207}Pb	$9.849 \times 10^{-10} \text{ yr}^{-1}$
^{146}Sm	^{142}Nd	$6.73 \times 10^{-10} \text{ yr}^{-1}$
^{40}K	^{40}Ar	$5.50 \times 10^{-10} \text{ yr}^{-1}$
^{40}K	^{40}Ca	$4.96 \times 10^{-10} \text{ yr}^{-1}$
^{238}U	^{206}Pb	$1.551 \times 10^{-10} \text{ yr}^{-1}$
^{130}Te	^{130}Xe	$8.66 \times 10^{-22} \text{ yr}^{-1}$
^{87}Rb	^{87}Sr	$1.42 \times 10^{-11} \text{ yr}^{-1}$
^{187}Re	^{187}Os	$1.64 \times 10^{-11} \text{ yr}^{-1}$
^{176}Lu	^{176}Hf	$1.93 \times 10^{-11} \text{ yr}^{-1}$
^{232}Th	^{208}Pb	$4.95 \times 10^{-11} \text{ yr}^{-1}$
^{40}K	^{40}Ar	$5.81 \times 10^{-11} \text{ yr}^{-1}$
^{147}Sm	^{143}Nd	$6.54 \times 10^{-12} \text{ yr}^{-1}$
^{138}La	^{138}Ce	$2.24 \times 10^{-12} \text{ yr}^{-1}$
^{130}Te	^{130}Xe	$8.66 \times 10^{-22} \text{ yr}^{-1}$

3. Temperature Range of Magma Emplacement

Granitic magma	700–800 °C
Basaltic magma	1250–1350 °C
Komatiitic magma	1650 °C

4. Average Compositions of Earth Shells

	$N =$ Atomic number	$M =$ Atomic mass (g/mol)	Unit	C1 Chondrite	Bulk Solid Earth
H	1	1.00794	%	2.02	0.0257
He	2	4.002602		ppm	56 (<i>nL/g</i>)
Li	3	6.941	ppm	1.49	1.08
Be	4	9.012182	ppm	0.0249	0.046
B	5	10.811	ppm	0.87	0.203
C	6	12.011	%	3.5	0.73
N	7	14.00674	ppm	3180	55
O	8	15.9994	%	46.4	30.5
F	9	18.998403	ppm	58.2	17.1
Ne	10	20.1797	ppm	203 (<i>pL/g</i>)	
Na	11	22.989768	%	0.51	0.18
Mg	12	24.3050	%	9.65	15.39
Al	13	26.981539	%	0.86	1.59
Si	14	28.0855	%	10.65	16.5
P	15	30.973762	ppm	1080	1101
S	16	32.066	%	5.4	0.634
Cl	17	35.4527	ppm	680	74.5
Ar	18	39.948	ppm	751 (<i>pL/g</i>)	
K	19	39.0983	%	0.0550	0.0164
Ca	20	40.078	%	0.925	1.71
Sc	21	44.95591	ppm	5.92	10.93
Ti	22	47.88	%	0.044	0.0813
V	23	50.9415	ppm	56	94.4
Cr	24	51.9961	ppm	2650	3720
Mn	25	54.93805	ppm	1920	1974
Fe	26	55.847	%	18.1	30.3
Co	27	58.9332	ppm	500	838
Ni	28	58.6934	ppm	10500	17 690
Cu	29	63.546	ppm	120	60.9
Zn	30	65.39	ppm	310	37.1
Ga	31	69.723	ppm	9.2	2.74

Table showing the compositions of the different Earth shells. Data are from Anders and Grevesse (1989) Taylor and Mc Lennan, Mc Donough et al. (1992), Mc Donough and Sun (1995), Rudnick and Fountain (1995), McDonough (1998), Rudnick and Gao (2003). Several other data are coming from the Geochemical Earth Reference Model (GERM) Internet site '<http://www.earthref.org>'.

(continued)

Core	Primitive Mantle	Upper Mantle	Oceanic Crust	Bulk Contin- ental Crust	Ocean	Atmos- phere
0.06	0.01	0.00025		0.14	10.8	0.000053
					0.0000069	5.2
0	1.6	1.57	10	16	0.18	
0	0.068	0.0442	0.5	1.9	0.000225	
0	0.3	0.07	4	11	4.39	
0.2	0.0120	0.0018		0.02	0.0028	0.0337
170	2	0.05		56	8.5	780900
0	44	44		45	85.7	20.95
0	25	9.8		553	1.3	
				0.00007	0.04	18
0	0.267	0.214	2.08	2.36	1.05	
0	22.8	22.8	4.64	3.20	0.135	
0	2.35	2.23	8.47	8.41	0.00000016	
6.4	21	21	23.1	26.77	0.0003	
3200	90	54	873	2000	0.07	
1.9	0.025	0.024		0.0404	0.0885	0.0001
200	17	2.55		244	19000	0.001
				1.2	0.6	9300
0	0.024	0.0024	0.125	0.99	0.038	
0	2.53	2.4	8.08	5.29	0.04	
0	16.2	15.4	38	21.9	0.0000006	
0	0.12	0.0928	0.9	0.54	0.0000001	
120	82	82	250	138	0.0019	
9000	2625	2625	270	135	0.0003	
4000	1045	1045	1000	852	0.0002	
85	6.26	6.26	8.16	5.6	0.00000056	
2500	105	105	47	26.6	0.00039	
52 000	1960	1960	135	59	0.0054	
125	30	29.1	86	27	0.0002	
0	55	55	85	72	0.0037	
0	4	3.8	17	16	0.00003	

(continued)

	N = Atomic number	M = Atomic mass (g/mol)	Unit	C1 Chondrite	Bulk Solid Earth
Ge	32	72.61	ppm	31	7.24
As	33	74.92159	ppm	1.85	1.67
Se	34	78.96	ppb	21 000	2560
Br	35	79.904	ppb	3570	300
Kr	36	83.80	ppb	8.7 (<i>pL/g</i>)	
Rb	37	85.4678	ppm	2.3	0.405
Sr	38	87.62	ppm	7.25	13.4
Y	39	88.90585	ppm	1.57	2.95
Zr	40	91.224	ppm	3.82	7.1
Nb	41	92.90638	ppm	0.240	0.45
Mo	42	95.94	ppm	0.9	1.66
Tc	43	98.9063	ppm		
Ru	44	101.07	ppb	710	1310
Rh	45	102.9055	ppb	130	242
Pd	46	106.42	ppb	550	1010
Ag	47	107.8682	ppb	200	54.4
Cd	48	112.411	ppb	710	76
In	49	114.82	ppb	80	7.5
Sn	50	118.71	ppb	1650	251
Sb	51	121.757	ppb	140	46.2
Te	52	127.60	ppb	2330	285.7
I	53	126.90447	ppb	450	50
Xe	54	131.29	ppb	8.6 (<i>pL/g</i>)	
Cs	55	132.90543	ppb	190	35
Ba	56	137.327	ppm	2.41	4.46
La	57	138.9055	ppm	0.237	0.437
Ce	58	140.115	ppm	0.613	1.148
Pr	59	140.90765	ppm	0.0928	0.171
Nd	60	144.24	ppm	0.457	0.844
Pm	61	146.9151	ppm		
Sm	62	150.36	ppm	0.148	0.274
Eu	63	151.965	ppm	0.0563	0.104
Gd	64	157.25	ppm	0.199	0.367
Tb	65	158.92534	ppm	0.0361	0.068
Dy	66	162.50	ppm	0.246	0.455
Ho	67	164.93032	ppm	0.546	0.102

(continued)

Core	Primitive Mantle	Upper Mantle	Oceanic Crust	Bulk Contin- ental Crust	Ocean	Atmos- phere
20	1.1	1.1	1.5	1.3	0.000051	
5	0.05	0.01	1	2.5	0.0026	
8000	75	71	160	130	0.00012	
700	50	5		880	65 000	
					0.29	1140
0	0.6	0.041	2.2	49	0.124	
0	19.9	12.93	130	320	8.1	
0	4.3	3.65	32	19	0.000013	
0	10.5	6.19	80	132	0.000026	
0	0.658	0.11	2.2	8	0.000015	
5	0.05	0.03	1	0.8	0.01	
4000	4.9	4.9	1	0.57	0.0007	
740	0.91	0.91	0.2	0.06		
3100	3.9	3.9	0.2	1.5		
150	8	7.8	26	56	0.28	
150	40	39.2	130	80	0.00011	
0	11	8.5	72	52	0.000115	
500	130	97.5	1400	1700	0.81	
130	5.5	1.1	17	200	0.33	
850	12	12	3	5		
130	10	1		710	64	
					0.066	86
65	21	504	30	2600	0.5	
0	6.6	0.45	25	456	0.021	
0	0.648	0.08	3.7	20	0.0000029	
0	1.675	0.538	11.5	43	0.0000028	
0	0.254	0.114	1.8	4.9	0.00000064	
0	1.25	0.738	10	20	0.0000028	
					—	
0	0.406	0.305	3.3	3.9	0.00000045	
0	0.154	0.119	1.3	1.1	0.0000013	
0	0.544	0.430	4.6	3.7	0.0000007	
0	0.099	0.080	0.87	0.6	0.00000014	
0	0.674	0.559	5.7	3.6	0.00000091	
0	0.149	0.127	1.3	0.77	0.00000022	

(continued)

	<i>N</i> = Atomic number	<i>M</i> = Atomic mass (g/mol)	Unit	C1 Chondrite	Bulk Solid Earth
Er	68	167.26	ppm	0.16	0.296
Tm	69	168.93421	ppm	0.0247	0.046
Yb	70	173.04	ppm	0.161	0.297
Lu	71	174.967	ppm	0.0246	0.046
Hf	72	178.49	ppm	0.103	0.191
Ta	73	180.9479	ppm	0.0136	0.025
W	74	183.85	ppm	0.093	0.170
Re	75	186.207	ppb	40	75.3
Os	76	190.2	ppb	490	900
Ir	77	192.22	ppb	455	835
Pt	78	195.08	ppb	1010	1866
Au	79	196.96654	ppb	140	164
Hg	80	200.59	ppb	300	23.1
Tl	81	204.3833	ppb	140	12.2
Pb	82	207.2	ppb	2470	232
Bi	83	208.98037	ppb	110	9.85
Po	84	208.9824			
At	85	209.9871			
Rn	86	222.0176			
Fr	87	223.0197			
Ra	88	226.0254			
Ac	89	227.0278			
Th	90	232.0381	ppm	0.029	0.054
Pa	91	231.03588			
U	92	238.0289	ppm	0.0074	0.0139

References

- Anders, E. and Grevesse, N., 1989. Abundances of the elements: Meteoritic and solar. *Geochimica Cosmochimica Acta*, 53: 197–214.
- McDonough, W.F., 1998. Earth's core. In: C.P. Marshall and R.W. Fairbridge (Editors), *Encyclopedia of geochemistry*. Kluwer Academic Publishers, Dordrecht, pp. 151–156.
- McDonough, W.F. and Sun, S.-S., 1995. Composition of the Earth. *Chemical Geology*, 120: 223–253.

(continued)

Core	Primitive Mantle	Upper Mantle	Oceanic Crust	Bulk Contin- ental Crust	Ocean	Atmos- phere
0	0.438	0.381	3.7	2.1	0.00000087	
0	0.068	0.060	0.54	0.28	0.00000017	
0	0.441	0.392	5.1	1.9	0.00000082	
0	0.0675	0.061	0.56	0.30	0.00000015	
0	0.283	0.167	2.5	3.7	0.0000016	
0	0.037	0.006	0.3	0.7	0.0000025	
0.47	0.029	0.002	0.5	1	0.000001	
230	0.28	0.27	0.9	0.188	0.0084	
2750	3.43	3.43	0.004	0.041	0.0000017	
2550	3.18	3.18	0.02	0.037	0.00000115	
5700	7.07	7.07	2.3	1.5	0.000117	
500	0.98	0.96	0.23	1.3	0.011	
50	10	9.8	20	30	0.15	
30	3.5	3.5	12	500	0.0123	
4000	150	20	800	11 000	0.03	
25	2.5	0.5	7	18	0.017	
0	0.0795	0.06	0.22	5.6	0.0000004	
0	0.0203	0.002	0.1	1.3	0.0033	

McDonough, W.F., Sun, S.-S., Ringwood, A.E., Jagoutz, E. and Hofmann, A.W., 1992.

Potassium, rubidium, and cesium in the Earth and Moon and the evolution of the mantle of the Earth. *Geochimica et Cosmochimica Acta*, 56(3): 1001–1012.

Rudnick, R.L. and Fountain, D.M., 1995. Nature and composition of the continental crust: a lower crustal perspective. *Reviews in Geophysics*, 33: 267–309.

Rudnick, R.L. and Gao, S., 2003. The Composition of the Continental Crust. In: R.L. Rudnick (Editor), *The Crust. Treatise on Geochemistry*. Elsevier-Pergamon, Oxford, pp. 1–64.

Taylor, S.R. and McLennan, S.M., 1985. *The continental crust: its composition and evolution*. Blackwell scientific Publications, Oxford, 312 pp.

5. Geological Time Scale and Live Evolution

Aeon	Era		System	Epoch	Ma	Life evolution (Ma)
PHANEROZOIC	Cenozoic	Quaternary	Quaternary	Holocene	Present	Modern human beings (0,12)
				Pleistocene	0,01	
		Tertiary	Neogene	Pliocene	1,75	Australopithecus (4)
				Miocene	5,3	
			Paleogene	Oligocene	23,5	
				Eocene	33,7	Mammal's expansion (50)
	Paleocene	53	1 st Primates (58)			
	Mesozoic	Secondary	Cretaceous		65	Dinosaur's extinction (65)
				Jurassic	135	
				Triassic	203	1 st Mammals (180)
	Paleozoic	Primary	Permian		250	1 st Dinosaurs (230)
				Carboniferous	295	
			Devonian		355	1 st Reptiles (360)
				Silurian	410	
			Ordovician		435	1 st Amphibians (440)
				Cambrian	500	1 st Aerial plants (450) 1 st Vertebrates (fishes) (500)
	PRECAMBRIAN	PROTEROZOIC			540	1 st Shelled Invertebrates (560) 1 st Pluricellular organisms (1000) 1 st Eukariotes (2000-2100)
			ARCHAEOAN		2500	1 st Cyanobacterias (2600-2700)
HADAEOAN			4000	1 st Bacterias (3500)		
				4550		

Simplified geological time scale for Earth; main stages in life development are also shown. The International Stratigraphic Chart is available on UNESCO-IUGS Internet site: ftp://ftp.iugs.org/pub/iugs/iugs_intstratchart.pdf

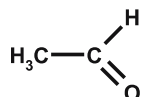
2.4 Biochemical Data

1. The Main Prebiotic Precursors of Biomolecules (in brackets, the IUPAC nomenclature recommendation)

water (hydrogen oxyde)



acetaldehyde (ethanal)



acetonitrile (methyl cyanide)



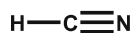
acetylene (ethyne)



ammonia



cyanhydric acid (hydrogen cyanide)



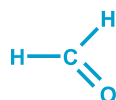
cyanoacetylene (2-Propynenitrile)



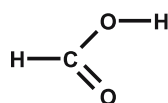
ethylene (ethene)



formaldehyde (methanal)



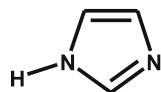
formic acid (methanoic acid)



hydrogen sulfide



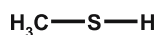
imidazole



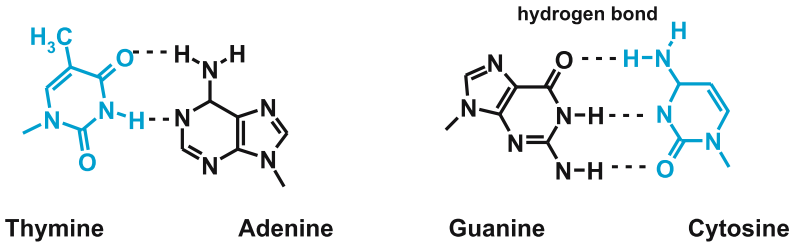
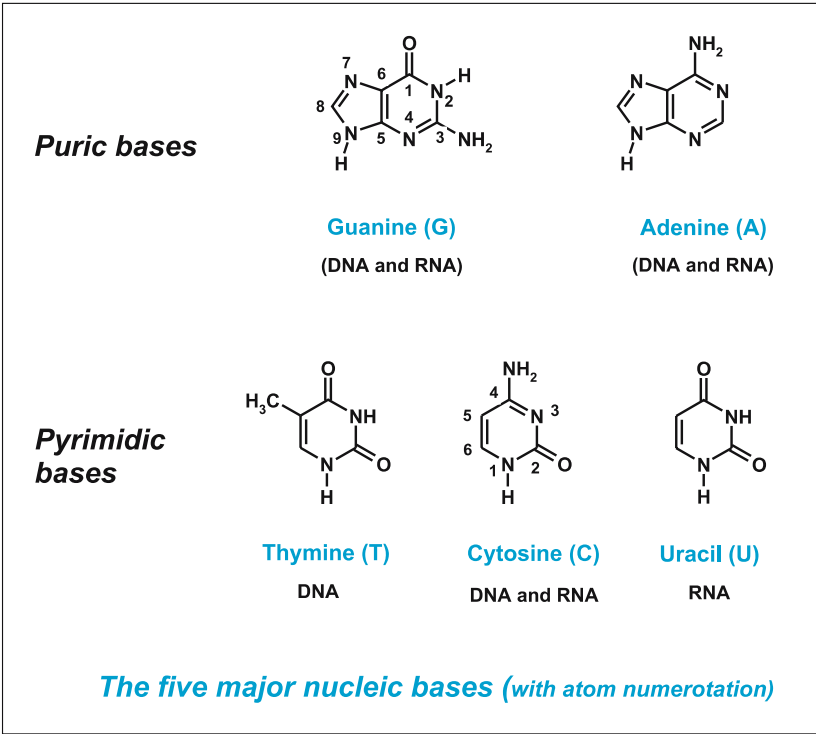
methane



methanethiol

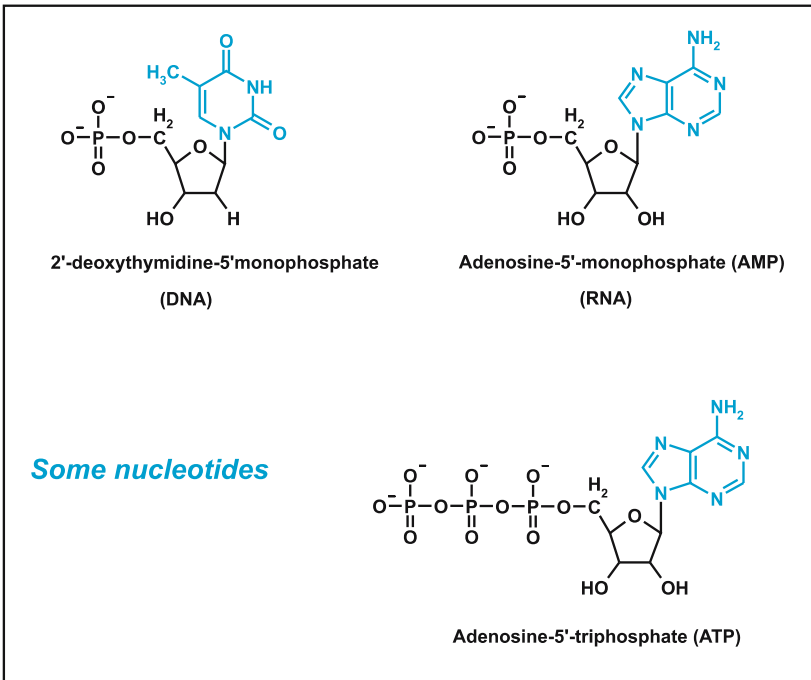
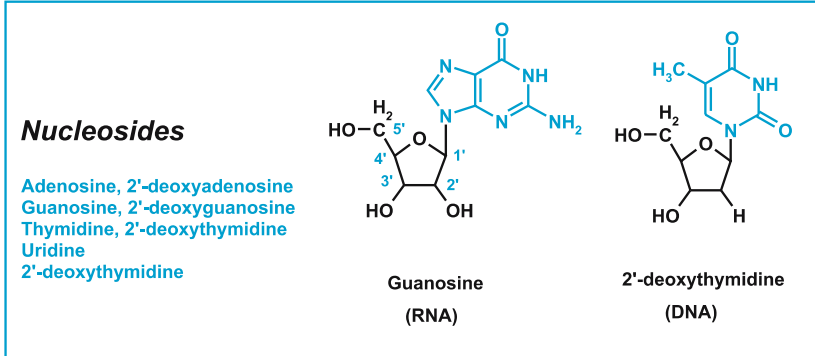


2. Nucleic Bases

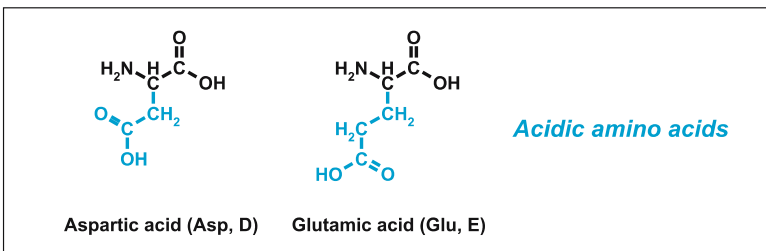
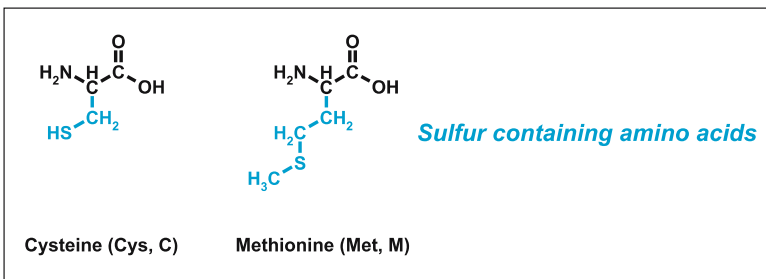
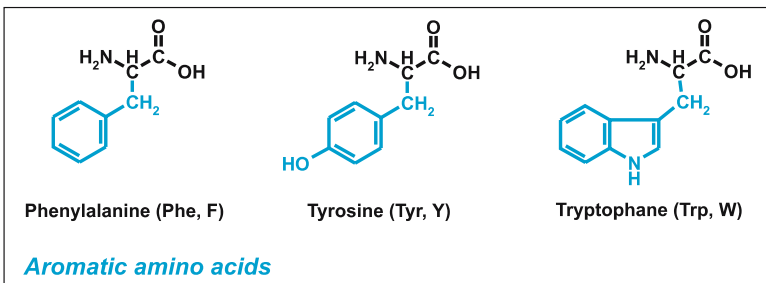
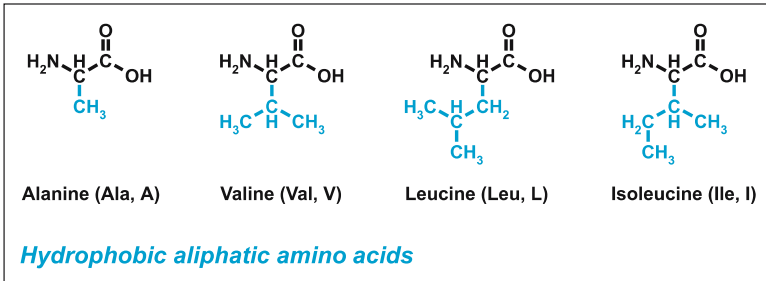


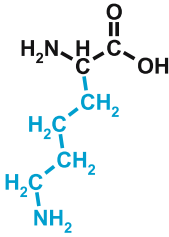
Watson-Crick base pairs

3. Nucleosides and Nucleotides

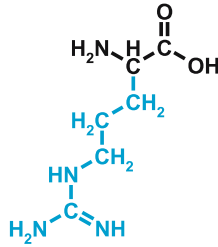


4. Amino Acids Genetically Coded

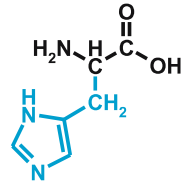




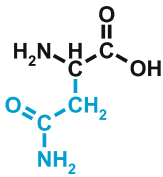
Lysine (Lys, K)



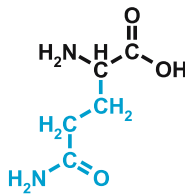
Arginine (Arg, R)



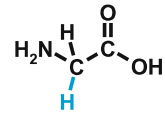
Histidine (His, H)

Basic amino acids

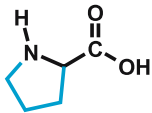
Asparagine (Asn, N)



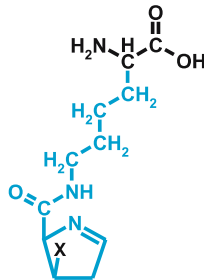
Glutamine (Gln, Q)



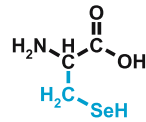
Glycine (Gly, G)



Proline (Pro, P)

N-alkyl amino acid

Pyrrolysine

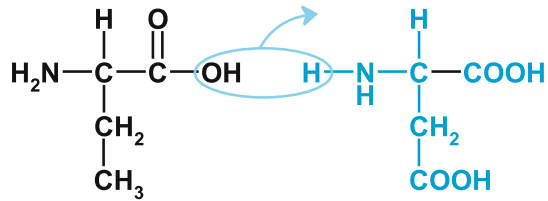
X = NH₂, CH₃ or OH

Selenocysteine

Rare amino acids

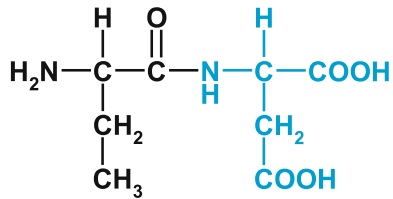
5. Peptide Bond Formation

leucine + aspartic acid \longrightarrow peptide + water



Leucine

Aspartic acid



Dipeptide Leu-Asp

3 Glossary

The field of astrobiology is by definition multidisciplinary, therefore, any glossary devoted to it is likely to be incomplete. Nevertheless, the editors hope that the following glossary, which contains approximately 1000 terms commonly used in astrobiology, will help specialists in one field to obtain pertinent information or clarifications about other disciplines.

This is an updated version of the glossary found in Lectures in Astrobiology, Vol. 1 of 2005.

M. Gargaud, Ph. Claeys, H. Martin

A

a (*astronomy*): Orbital parameter, semi-major axis of elliptical orbit.

Abiotic: In absence of life.

Ablation (*astronomy*): Loss of matter by fusion or vaporization during the entrance of an object into the atmosphere.

Absolute magnitude: Magnitude of a stellar object when seen from a distance of 10 parsecs (32.6 light-years) from Earth.

Acasta: Region of Northern Territories (Canada) where the oldest relics of continental crust were discovered. These gneisses were dated at 4.03Ga.

Accretion (*astronomy*): Matter aggregation leading to the formation of larger objects (stars, planets, comets, asteroids). Traces of accretion of their parent body (asteroid) can be observed in some meteorites.

Accretion disk: Disk of matter around a star (or around a black hole) such that matter is attracted by the central object and contributes to its growth. Disks around protostars, new-born stars and T Tauri stars are accretion disks.

Accretion rate (*astronomy*): Mass accreted per time unit (typically 10^{-5} to 10^{-8} solar mass per year).

Accuracy: The accuracy of a method is its capacity to measure the exact value of a given quantity.

Acetaldehyde: CH₃CHO, ethanal.

Acetic acid: CH_3COOH , ethanoic acid.

Acetonitrile: CH_3CN , cyanomethane, methyl cyanide.

Achiral: Not chiral.

Acidophile: Organism, which *likes* acidic media, which needs an acidic medium.

Actinolite: Mineral, Inosilicate (double chain silicate), $[\text{Ca}_2(\text{Fe},\text{Mg})_5 \text{Si}_8\text{O}_{22}(\text{OH},\text{F})_2]$. It belongs to the amphibole group.

Activation energy (E_a): Empirical parameter that describes, via Arrhenius law, the temperature dependence of a reaction rate. In the case of a reaction $\text{A} + \text{B} \rightarrow \text{C}$, E_a can be described as the energy difference between the reactants $\text{A} + \text{B}$ and the activated complex (AB^*) on the reaction pathway to C . It corresponds to an energy barrier.

Active site (of an enzyme): Part of an enzyme where substrate is specifically fixed, in a covalent or non-covalent way, and where catalysis takes place.

Activity (solar or stellar): All physical phenomena that are time dependent and related to star life (like stellar wind, solar prominence, sun spots). Their origin is mainly magnetic, they correspond to emissions of electromagnetic waves at different frequencies (from radio waves to X-rays) and also to emissions of charged particles (protons, alpha particles and heavier particles). In the case of the Sun, the solar activity is cyclic. The shorter cycle is eleven years.

Activity: In a non-ideal solution, activity plays the same role as mole fraction for ideal solution.

Adakite: Volcanic felsic rock generated in subduction zone, by partial melting of the subducted basaltic oceanic crust.

Adaptive optics: The goal of adaptive optics is to correct the effects of atmospheric turbulence on the wavefronts coming from stars. The basic idea consists of measuring the wavefront form each 10ms (to follow the atmospheric turbulence), and to deform the mirror(s) accordingly to correct and stabilize the wavefront.

Adenine: Purine derivative which plays an important role in the living world as component of nucleotides.

Adenosine: Nucleoside that results from condensation of adenine with ribose.

ADP: Adenosine diphosphate (see also ATP).

Aerobian: Organism whose life requires the presence of free oxygen in its environment.

Aerobic respiration: Ensemble of reactions providing energy to a cell, oxygen being the ultimate oxidant of organic or inorganic electrons donors.

Aerosols: Liquid or solid submillimeter particles in suspension in a gas. Aerosols play an important role in atmospheric physics and chemistry.

Affinity constant: The efficiency of recognition of a target by a ligand is characterized by an affinity constant (in M^{-1}), which is the ratio of ligand-target concentration ($[\text{ligand-target}]_{\text{eq}}$) to the product of the free ligand and free target concentrations at equilibrium ($[\text{ligand}]_{\text{eq}} \times [\text{target}]_{\text{eq}}$): $K = ([\text{ligand-target}]_{\text{eq}} / [\text{ligand}]_{\text{eq}} \times [\text{target}]_{\text{eq}})$.

AIB: See aminoisobutyric acid.

Akilia: Region of Greenland near Isua where there are exposed sediments and volcanic rocks similar in both composition and age (3.865 Ga) to the Isua gneisses.

Alanine (*Ala*): Proteinic amino acid containing three carbon atoms.

Albedo: Fraction of the incident light that is reflected by a reflecting medium (i.e., atmosphere) or surface (i.e., ice cap). A total reflection corresponds to an albedo of 1.

Albite: Mineral. Tectosilicate (3D silicates). Sodic plagioclase feldspar $\text{NaAlSi}_3\text{O}_8$.

Alcalophile: Organism, which *likes* alkaline (basic) media, or needs an alkaline medium.

Alcohol: R-OH.

Aldehyde: R-CHO.

Aldose: Any monosaccharide that contains an aldehyde group (-CHO).

ALH84001: Martian meteorite found, in 1984, in the Alan Hills region (Antarctica). In 1996, the unverified claim that it contains traces of metabolic activity and even, possibly, microfossils was the starting point of strong debates.

Aliphatic hydrocarbon: Hydrocarbon having an open chain of carbon atoms. This chain can be normal or forked.

Allende: Large carbonaceous chondrite (meteorite) of 2 tons, of the C3/CV type and found in Mexico in 1969.

Allochthonous sediment: Sediment that formed in a place different of its present day location (transported).

Alpha helix: A type of secondary helical structure frequently found in proteins. One helix step contains approximately 3.6 amino acid residues.

Alteration (*Weathering*): Modification of physical and chemical properties of rocks and minerals by atmospheric agents.

Amide: R-CO-NH₂, R-CO-NH-R', R-CO-NR'R'' depending if the amide is primary, secondary or tertiary. The bond between the CO group and the N atom is generally called the amide bond but when it links two amino acid residues in a polypeptide or in a protein, it is called the peptide bond.

Amine: Derivatives of ammonia NH_3 in which one, two or three H atoms are substituted by an R group (a group containing only C and H atoms) to give primary, secondary or tertiary amines.

Amino acid (AA): Organic molecule containing a carboxylic acid function (COOH) and an amino function (generally but not always a NH_2 group). If the two functions are linked to the same carbon atom, the AA is an alpha amino acid. All proteinic AA are alpha amino acids.

Amino acids (biological): AA directly extracted from organisms, fossils, sedimentary rocks or obtained, after hydrolysis, from polypeptides or proteins found in the same sources.

Amino acids (proteinic): AA found as building blocks of proteins. All proteinic AA are homochiral and characterized by an L absolute configuration.

Amino isobutyric acid (AIB): Non-proteinic alpha amino, alpha methyl AA. Detected in chondrites.

Amino nitrile: Molecule containing a CN group and an amine function. It could have played a role in the prebiotic synthesis of amino acids.

Amitsôq: Region of Greenland where the oldest huge outcrops ($\sim 3000 \text{ km}^2$) of continental crust, dated at 3.822 Ga (gneiss) are exposed.

Amorphous: Solid state characterized by a lack of order at large distances.

Amphibole: Mineral family, Inosilicate (water-bearing double chain silicate), including actinolite, hornblende, glaucophane, etc.

Amphibolite: Rock generated by metamorphism of basalt. It mainly consists in amphibole and plagioclase feldspar crystals, sometimes associated with garnet.

Amphiphile: Molecule with a hydrophilic part and a lipophilic part.

Amplification (of DNA): Production, in relatively large quantity, of fragments of DNA by in vitro replication, starting from a very small initial sample (see PCR).

Amu (atomic mass unit): Atomic mass unit such that the atomic mass of the ^{12}C (carbon isotope) has exactly a mass equal to 12.0000.

Anabolism: General term to designate a group of biochemical reactions involved in the biosynthesis of different components of living organisms.

Anaerobian: Organism, which does not need free oxygen for its metabolism. In some cases, free oxygen is a poison for anaerobian organisms.

Anaerobic respiration: Ensemble of reactions providing energy to a cell, the ultimate oxidant being an inorganic molecule other than oxygen.

Anatexis: High degree metamorphism where rock begins to undergo partial melting.

Andesite: Effusive magmatic rock (volcanic); it mainly consists in sodic plagioclase feldspar + amphibole + pyroxene crystals. These magmas are abundant in subduction zones. Diorite is its plutonic equivalent.

Angular momentum: The angular momentum, L of a rigid body with a moment of inertia I rotating with an angular velocity ω is $L = I \times \omega$. In the absence of external torque, the angular momentum of a rotating rigid body is conserved. This constancy of angular momentum allows to fix constraints on star and planetary system genesis from huge interstellar gas clouds.

Anorthite: Mineral, Tectosilicate (3D silicates), Calcic plagioclase feldspar $\text{CaAl}_2\text{Si}_2\text{O}_8$.

Anorthosite: Plutonic magmatic rock only made up of plagioclase feldspar. Generally, it consists in an accumulation (by flotation) of plagioclase crystals. Relatively rare on Earth, anorthosites are widespread on the Moon where they form the primitive crust.

Anoxic: Environment (water, sediment, etc.) where oxygen is highly deficient or absent. Synonym: anaerobic.

Anthophyllite: Mineral, Inosilicate (double-chain silicate), $[(\text{Mg})_7\text{Si}_8\text{O}_{22}(\text{OH},\text{F})_2]$. It belongs to the amphibole group.

Antibody (immunoglobulin): Glycoprotein produced because of the introduction of an antigen into the body, and that possesses the remarkable ability to combine with the very antigen that triggered its production. Polyclonal antibodies are produced by different types of cells (clones), whereas monoclonal antibodies are exclusively produced by one type of cell by fusion between this cell and a cancerous cell.

Anticodon: Triplet of nucleotides of tRNA able to selectively recognize a triplet of nucleotides of the mRNA (codon).

Antigen: Foreign substance that in an organism, is able to induce the production of an antibody.

Antisense: Strand of DNA that is transcribed into a mRNA (messenger RNA).

Aphelia: In the case of an object in elliptical motion around a star, the point that corresponds to the largest distance with respect to the star.

Apollo: Ensemble of asteroids whose orbits intersect that of the Earth.

Aptamer: Synthetic polynucleotide molecule that binds to specific molecular targets, such as a protein or oligonucleotide. Contrary to nucleic acids, aptamers are able to fold back into a 3D structure, thus leading to the formation of binding cavities specific to the target.

Archaea: One of the three main domains of the living world. All the organisms of this domain are prokaryotes. Initially, these organisms were considered as the most primitive form of life but this view is no longer accepted by most biologists.

Most extremophiles such as hyperthermophiles and hyperacidophiles belong to the archaea domain.

Archaean (*Eon*): Period of time (eon) ranging from 4.0 to 2.5Ga. The Archaean eon belongs to Precambrian. Unicellular life existed and possibly was already aerobic.

Archeobacteria: See Archaea.

Arginine: Proteinic alpha amino acid containing six carbon atoms and a guanido group in the side chain.

Aromatic hydrocarbon: Hydrocarbon that contains one or several benzene ring(s).

Aspartic acid (*Asp*): Proteinic AA with an acidic side chain.

Assimilative metabolism: In a cell, reduction process by which an inorganic compound (C, S, N, etc.) is reduced for use as a nutriment source. The reduction reaction necessitates an energetic supply (endothermic).

Asteroid belt: Ring-shaped belt between Mars and Jupiter where the majority of the asteroids of the solar system are located.

Asteroid: Small object of the solar system with a diameter less than 1000km. Many of them are orbiting around the Sun, between Mars and Jupiter (asteroid belt).

Asthenosphere: Layer of the Earth mantle, located under the lithosphere and having a ductile behavior. It is affected by convective movements. Depending on the geothermal gradient, its upper limit varies between 0km under mid-oceanic ridges and 250km under continents.

Astrometry: Part of astrophysics that studies and measures the position and motion of objects in the sky. By extension, astrometry refers to all astrophysics phenomena that can be documented or studied by observation and measurement of the motion of celestial object.

Asymmetric atom: Synonym for chirotopic carbon. See asymmetric.

Asymmetric: An atom is called asymmetric when it is surrounded by four ligands, which are oriented in 3D space in such a way that they define an irregular tetrahedron. Such an atomic arrangement is chiral and can exist as two enantiomeric forms. These two enantiomers are called D or L depending on their absolute configuration. The D/L nomenclature is now replaced by the R/S nomenclature of Cahn, Ingold and Prelog but the D/L system is still accepted for amino acids and oses. The asymmetric carbon atom is a particular (but very important) example of asymmetric atom.

Atmosphere: Gaseous envelope around a star, a planet or a satellite. In absence of rigid crust or ocean, the atmosphere is defined as the most external part of the object. *The primary atmosphere* of a young planet corresponds to the first gaseous envelope directly generated from protostellar nebula. *The primitive*

atmosphere of the Earth corresponds to the atmosphere in which prebiotic chemistry occurred, when free oxygen was in very low concentrations. In the second part of the past century, this primitive atmosphere was considered as being highly reductive. Today, it is considered as mainly consisting of carbon dioxide, nitrogen and water vapor.

ATP synthetase: Enzyme involved in ATP synthesis.

ATP: Adenosine triphosphate. Molecule that plays an important role in the living world for the energy transfers. Its hydrolysis in ADP and in inorganic phosphate is an exergonic reaction.

AU (*astronomical unit*): Average Earth–Sun distance corresponding to 149.6×10^6 km (or approximately 8 minute light and 100 Sun diameters).

Authigenic minerals: In a sedimentary environment, minerals generated by direct local precipitation of dissolved ions.

Autocatalysis: Chemical reaction such that a reaction product acts as a catalyst for its own synthesis.

Autochthonous sediment: Sediment that formed in the place where it is now.

Automaton (*chemical automaton*): As defined by Brack, chemical system able to promote its own synthesis.

Autotroph: Organism, which is able to synthesize its own constituents from simple molecules like water and carbon dioxide.

B

Bacteria: One of the three domains of the living world. Organisms of this domain are all prokaryotes. They are the more abundant microorganisms on Earth. One of their characteristics consists in the presence of a cell wall containing muramic acid. Some of these organisms are extremophiles like *Aquifex* and *Thermoga*.

Bacteriochlorophylls: Pigment found in microorganisms and containing a tetrapyrrole acting as a ligand for an Mg cation (similarly to chlorophylls of green plants).

Barophile: Microorganism that lives optimally (or can only grow) in high pressure environments such as deep sea environments.

Basalt: Effusive mafic magmatic rock (volcanic); it mainly consists in plagioclase feldspar + pyroxenes \pm olivine crystals. It results of 20 to 25% melting of the mantle. Gabbro is its plutonic equivalent.

Bases in nucleotides: See nucleic base, purine bases and pyrimidine bases.

Benioff plane: In a subduction zone, interface between the subducted slab and the overlying mantle wedge.

Beta sheet: A secondary pleated structure frequently observed in proteins.

BIF (*Banded Iron Formation*): Sedimentary rocks widespread in Archaean terrains and no longer generated today. They consist in alternation of black iron-rich (magnetite) and white amorphous silica-rich layers.

Bifurcation (*thermodynamics*): This term describes the behavior of a system, which submitted to a very small variation in exchange conditions with its surroundings, jumps suddenly from one stationary state to another one. The characteristics of new stationary state cannot be predicted on the basis of a complete knowledge of the initial stationary state.

Binding energy (*nuclear*): In an atom the nuclear binding energy is the energy necessary to break the atomic nucleus into its elementary components. This is also the amount of energy released during the formation of an atomic nucleus.

Biocenosis: Group of interacting organisms living in a particular habitat and forming an ecological community (ecosystem).

Biochemical sediment: Sediment formed from chemical elements extracted from water by living organisms (for instance, limestone formed by an accumulation of calcium carbonate shells).

Biofilm: Tiny film (few μm thick) consisting in colonies of microorganisms fixed on an inorganic or organic surface.

Biogenic sediment: Sediment formed by precipitation of ions dissolved in water by the mean of living beings (e.g., calcium ion in shells).

BIOPAN: Experimental module made by ESA to be fixed on a Russian satellite of the Photon type, the aim of the system is to expose samples or dosimeters to space conditions as vacuum, microgravity or radiation.

Bioprecipitation: Precipitation of solid mineral phases linked to the metabolism of a living organism. Through selective evolution, this process was used by organisms to reduce the toxic excess of some soluble compounds or elements (for example Ca), then to constitute new functionalities (exo- or endoskeleton, protective tests, teethes, etc. In some extent, the organisms can favor bioprecipitation even when the surrounding physicochemical conditions are unfavorable. This is for example the case of diatoms that precipitate an SiO_2 test under seawater conditions, which would normally dissolve SiO_2 .

Biosensor: Device containing biological compounds on/in a membrane, which translates biological parameters such as electric potential, movement, chemical concentration, etc., into electrical signals.

Biosignature: Observable considered as an evidence of the presence of life.

Biosphere: Ensemble of species living on Earth.

Biosynthesis: Production of a chemical compound by a living organism, equivalent to anabolism.

Biotectonic: Environmental concept proposed to account for the situation in Lake Vostok (Antarctica), where crustal tectonics (faults) controls and activates hydrothermal circulation. The latter provides chemical elements allowing life to survive. In the ocean, as well as in the oceanic crust, life linked to magmatic activity should rather be called biomagmatic.

Biotite: Mineral, Phyllosilicate (water-bearing sheet silicate), $[\text{K}(\text{Fe},\text{Mg})_3\text{Si}_3\text{AlO}_{10}(\text{OH})_2]$. It belongs to the mica group and is also called black mica.

Biotope: Smallest unit of habitat where all environmental conditions and all type of organisms found within are the same.

Bipolar flow (*astronomy*): Flow of matter in two opposite directions, perpendicularly to the circumstellar disk associated to a new-born star. The components of the bipolar flow are molecules, atoms, ions and dust particles.

Birthline: Locus, in the Hertzsprung–Russell diagram where young stars become optically visible.

Black body (*radiation*): Radiation emitted by a body at a temperature T and such that the coupling between matter and radiation is perfect. Such a body is black. The total power emitted by unit area and the power emitted at a well-defined frequency depends only on the temperature (the T^4 dependence is given by the empirical Stefan law and by the theoretical Planck law). As a first approximation and on the basis of their emission properties, stars and planets can be described as black bodies.

Black smoker: Also called hydrothermal vent. Structure observed on the oceans floor generally associated with mid-ocean ridges. There, hot hydrothermal fluid, rich in base metal sulfides, comes in contact with cold oceanic water. Polymetallic sulfides and calcium sulfate precipitate progressively building a columnar chimney around the vent.

Blast: Blast (Basic Local Alignment Search Tool) represents a powerful and rapid way to compare new sequences to an already existing database, which may either contain nucleotides (Blastn) or proteins (Blastp). Since the BLAST algorithm establishes local as well as global alignments, regions of similarity embedded in otherwise unrelated proteins could be detected. The Blast program gives a value for each of the high scoring results, together with the probability to find an identical score in a database of the same size by chance.

Blue algae: See cyanobacteria. Old name for some procaryotic unicellular, which are not algae.

Bolometric light (*or bolometric magnitude*): Total radiation output by time unit of a stellar object.

Bootstrap: Statistical method with re-sampling, commonly used to measure the robustness and the reliability of phylogenetic trees.

Braking radiation: Electromagnetic radiation emitted by high speed particle (electron, proton) when deviated by a magnetic or electric field. (Also see Bremsstrahlung).

Branching ratio: In the case of a chemical or of a nuclear reaction, such that different reaction paths exist, the branching ratio is the relative rate constants or probabilities of occurrence per unit of time for each different path.

Breccias (*geology*): Sedimentary or magmatic rocks consisting in an accumulation of angular fragments in a sedimentary or magmatic matrix.

Bremsstrahlung: German word also used in French and English to describe the electromagnetic radiation emitted by high speed particle (electron, proton) when deviated by a magnetic field.

Brown dwarf: Space body born as a star, but whose mass is too small (<80 Jupiter mass) to allow nuclear reaction: both core temperature and core pressure are insufficient to initiate hydrogen fusion.

C

C (*alpha*): Carbon atom linked to a chemical function we are interested in. More specifically, the carbon atom directly linked to the carboxylic function in amino acids. In alpha amino acids, the amino group is linked to the alpha carbon. The Greek letters alpha, beta, gamma, etc., are used to describe carbon atoms separated from the function by one, two, three, etc., other carbon atoms.

CAI (*Ca-Al Inclusions*): Ca and Al-rich inclusions abundant in some chondrite meteorites.

Caldera: Circular kilometer-sized structure due to the collapse of superficial formation induced by the emptying of an underlying magma chamber.

Carbonaceous chondrite: Chondrite with high carbon content. The famous Murchison meteorite is a carbonaceous chondrite.

Carbonate compensation depth (*CCD*): Depth below which sedimentary carbonates are systematically dissolved in seawater. This depth depends on the physicochemical conditions of the environment (amount of sinking particles due to surface biological production, deep-water circulation, etc.). These conditions can change from one ocean to another and may also change through time. The present Atlantic ocean CCD is at about 3500m.

Carbonate: $(\text{CO}_3)^{2-}$ -bearing mineral, e.g., Calcite = CaCO_3 ; Dolomite = $\text{MgCa}(\text{CO}_3)_2$.

Carbonation: In Ca, Mg, K, Na and Fe-bearing minerals, chemical reaction of alteration resulting in the formation of carbonates.

Carbonic anhydride: CO_2 , synonym for carbon dioxide.

Carbonyl: $-(\text{CO})-$, this chemical function is found in carboxylic acids, ketones, aldehydes, amides and many other organic molecules.

Carboxylic acid: Organic molecule containing a COOH group.

C-asteroid: Asteroid containing carbon. C-asteroids are the parent-bodies of carbonaceous chondrites.

Catabolism: Part of metabolism involving all degradation chemical reactions in the living cell resulting in energy production. Large polymeric molecules such as polysaccharides, nucleic acids, and proteins are split into their smallest size constituents, such as amino acids, after which the monomers themselves can be broken down into simple cellular metabolites. The small size constituents can be subsequently used to construct new polymeric molecules. These reactions involve a process of oxidation that releases chemical free energy, part of which is stored through synthesis of adenosine triphosphate (ATP). All these reactions require enzymatic catalysis.

Catalysis: Chemical process such that a substance (catalyst) increases reaction rate by changing reaction pathway but without being chemically modified during reaction. Enzymes are very efficient catalysts able to increase reaction rates by several orders of magnitude and also able to limit the number of secondary products and therefore to increase reaction selectivity.

CCD (*Charge-Coupled Device*): Silicon photo electronic imaging device containing numerous photosensors (often at least 1000×1000). The most used astronomic detector in the visible wavelength domain.

Cell: Complex system surrounded by a semi-permeable membrane that can be considered as the basis unit of all living organisms.

Cenancestor: Last ancestor shared by all living beings. Synonym of LUCA.

Cenozoic (*Era*): Period of time ranging from 65Ma to present day. It corresponds to the Tertiary Era including the Quaternary.

Chalcophile: Chemical element frequently associated with sulfur (i.e., Cu, Zn, Cd, Hg, etc.).

Chandrasekhar mass: Also called Chandrasekhar limit. In astrophysics, the maximum possible mass of a white dwarf star. This maximum mass is about 1.44 times the mass of the Sun. After having burned all its nuclear fuel the star atmosphere collapses back on the core. If the star has a mass below the Chandrasekhar limit, its collapse is limited by electron degeneracy pressure; if the mass is above the Chandrasekhar limit it collapses and becomes a neutron star and possibly a black hole.

Chemical derivatization: Typically, derivatization consists in the chemical modification of a compound in order to produce a new compound easier to detect due to its new physicochemical properties (volatility, absorption in ultraviolet, fluorescence, etc.). The derivatization of a molecule involves altering part of it slightly or adding a new part to the original compound.

Chemical sediment: Sediment formed by direct precipitation of ions dissolved in water.

Chemolithoautotroph: Chemotroph that uses CO_2 as only source of carbon.

Chemolithotroph: Chemotroph that takes its energy from the oxidation of inorganic molecules such as NH_3 , H_2S , Fe^{++} . On Earth, the first living organisms could have been chemolithotrophs.

Chemoorganotroph: Chemotroph that takes its energy from oxidation of organic molecules.

Chemotroph: Organism that takes its free energy from the oxidation of chemicals.

Chert: Microcrystalline to cryptocrystalline sedimentary rock consisting of quartz crystals whose size is $<30\mu\text{m}$, or of amorphous silica.

Chicxulub: Large ($>180\text{km}$ in diameter) impact crater, located in the Gulf of Mexico. It is assumed to result of the collision of a big (10km in diameter) meteorite, 65Ma ago. This impact is considered as the cause of the important biological crisis at the Cretaceous-Tertiary boundary that led to mass extinction of many living species, such as ammonites, dinosaurs, etc.

Chirality: Property of an object (and therefore of a molecule) to be different from its mirror image in a plane mirror. A hand is an example of a chiral object (in Greek: *cheir* means hand). Any object is chiral or achiral and if it is chiral, it can exist as two enantiomorphous forms (called enantiomers for molecules).

Chlorophylls: Pigments of major importance in the oxygenic photosynthesis. The chemical structure of all chlorophylls is based on a porphyrin ring system chelating an Mg^{2+} cation. Chlorophylls are found in higher plants, algae and some microorganisms.

Chloroplast: Subcellular structure that plays a fundamental role for photosynthesis in all photosynthetic eukaryotes. Chloroplasts have more probably an endosymbiotic origin.

Chondre: Small spherical aggregate of radiated silicate minerals (typically 1mm in diameter), which is frequent in stony meteorites and especially in chondrites. Olivine is the main component of chondres.

Chondrite: Undifferentiated stony meteorite unmelted and frequently considered as a very primitive object. Chondrites have the same composition as the Sun except for volatile elements.

Chromatographic co-elution: Refers to organics compounds that cannot be separated by a given chromatographic method and that migrate at the same speed leading to a single detectable signal.

Chromatography: Preparative or analytical chromatography: experimental method based on the properties of all molecules to be absorbed more or less selectively by a solid phase (the stationary phase) and therefore to migrate at different rates when they are pushed away by a mobile phase, which can be a gas (gas chromatography or GC) or a liquid (LC or HPLC for High Performance LC).

Chromophore: Chemical group that absorbs UV and visible wavelengths, thus coloring the molecules or objects that contain them.

Chromosome: Subcellular structure containing most of the genetic material of the cell. Prokaryotic cells generally contain only one chromosome made of a circular DNA molecule while eukaryotic cells generally have several chromosomes, each of them containing a linear DNA molecule.

CIP (*Cahn-Ingold-Prelog*): General nomenclature used in chemistry to describe chiral molecules and more generally stereoisomers. Following the International Union of Pure and Applied Chemistry (IUPAC), CIP nomenclature must replace all other nomenclatures including D, L Fisher nomenclature except for amino acids and sugars for which the D/L nomenclature is still accepted.

Circular dichroism: The absorption coefficient of right and left circularly polarized light are different if the absorbing medium is chiral. By plotting the difference between the absorption coefficients as a function of the light wavelength, the curve that is obtained corresponds to the circular dichroism curve of the medium.

Circumstellar disk: Disk of gas and dust particles around a star.

Class [0,I,II,III] (*astronomy*): Classification of young stellar objects based on their electromagnetic emission in the microwave and infrared domains. It consists in an evolution sequence from the protostars (0 and I) to the T Tauri stars (III).

Clast: Fragment of mineral or rock included in another rock.

Clay: Mineral family. Phyllosilicate (water-bearing sheet silicate), e.g., kaolinite, illite, smectite, montmorillonite.

CM matter: Pristine material of the solar system, analogue to the constitutive matter of CM carbonaceous meteorites (M = Mighei) and very abundant in micrometeorites.

CNO cycle: Series of hydrogen burning reactions producing helium by using C, N, O as catalysts. These reactions provide the energy of the main sequence stars with a mass $> 1.1 M_{\text{Sun}}$.

Coacervat (*droplet*): Protein and polysaccharides containing emulsion. According to Oparin, model of protocells.

Codon: Triplet of nucleotides in a mRNA molecule that corresponds to a specific amino acid of a protein synthesized in a ribosome or to a punctuation signal in protein synthesis.

Coenzyme: Small molecule that binds with an enzyme and is necessary to its activity. ATP, CoA, NADH, biotine are examples of coenzymes. Coenzymes are often derived from vitamins.

Collapse (*astronomy*): Process that describes the formation of stars from dense cores. The process seems to be fast: less than 10^5 years.

Coma: Broadly spherical cloud of dust particles and gaseous molecules, atoms and ions surrounding cometary nucleus. It appears when nucleus becomes active, generally when approaching the Sun.

Combustion: Used to describe any exothermal chemical reaction involving dioxygen and organic reactants. Sometimes used in astrophysics to describe the thermonuclear fusion reactions taking place in the stars and, which, obviously, are also exothermal.

Comet: Small body of the solar system with an average size of 1 to 100km, traveling generally on a strongly elliptic orbit around the Sun. Comets are constituted of ice and dust and are considered as the most primitive objects of the solar system.

Cometary nucleus: Solid part of a comet (1–100km diameter), made of ice (H₂O, CH₃OH, CO, etc.) and of dust.

Cometary tail: Part of an active comet; three different cometary tails are known depending from their composition (dust particles and molecules, neutral sodium atoms, ions).

Complex molecule (*astronomy*): Molecule containing more than 3 atoms.

Complexation: Chemical term used to describe the non-covalent interaction of a molecule or, more frequently, an inorganic ion with other molecules (called ligands) to give a supramolecular system described as a complex.

Condensation: Chemical reaction involving two molecules and leading to the formation of a new chemical bond between the two subunits but also to the elimination of a small molecule (generally a water molecule). The formation of bonds between the subunits of many biochemical polymers are condensation reactions (examples: polynucleotides, polypeptides, polysaccharides).

Configuration: Term used in stereochemistry. Stereoisomers (except if they are conformers) have different configurations. As an example, butene can be cis or trans and it corresponds to two different configurations. In the case of chiral the D- and L-valine, the stereoisomers are enantiomers and they have opposite configurations. It is important to make a clear distinction between the relative configurations of two enantiomers and the absolute configuration of each of them.

Conglomerate (*chemistry*): In the restricted case of crystallized chiral molecules, a crystalline state such that all molecules of the same chirality (homochiral) crystallize together giving a mixture of crystal which, themselves, are of opposite chirality.

Conglomerate (*geology*): Detrital sedimentary rock consisting in an accumulation of large (>2mm) rounded fragments with or without a sedimentary matrix.

Continental margin: Submarine part of a continent making the boundary with oceanic crust.

Continents: Emerged and associated shallow depth (<300m) parts of the Earth surface. Their average composition is that of a granitoids.

Continuum (*astronomy*): In emission (or absorption) spectroscopy, emission (or absorption) background of a spectrum extending on a large frequency domain. Frequently, lines are superimposed on the continuous spectrum. Black body radiation corresponds to a continuous spectrum.

Cool Early Earth (*Model*): Period of time between Earth accretion (4.55 Ga) and the late heavy bombardment (4.0–3.8 Ga). This model considers that the early Earth was characterized by relatively cool temperature and low rate of meteoric impacts, thus being potentially favorable for life development.

Core (*geology*): Central shell of a planet. On Earth the core is mainly composed of iron with minor amounts of nickel and some traces of sulfur; it represents 16% in volume but 33% in mass of the planet. It is subdivided in a solid inner core from 5155 to 6378km depth and a liquid outer core from 2891 to 5155km depth.

Coronagraphy: Technique used in astronomy and that consists of masking one part of the observed field (for instance a bright star) to allow the observation of a lower luminosity object located in the immediate vicinity. This technique was first developed by Lyot, who masked the sun disk to observe its corona, thus leading to the given name of coronagraphy.

COROT: French project for the search of extrasolar planets based on a 25cm telescope and able to detect planets having a diameter equal to two times the Earth diameter and located at 0.5AU from its star.

Cosmic rays: Highly accelerated ions coming from the Sun (solar wind, essentially protons) or coming from other and extrasolar sources (galactic cosmic rays).

Covalent bonds: Interatomic bonds such that two atoms share one, two or three electrons pairs leading to the formation of a single, a double or a triple bond. A covalent bond is described as a polarized bond if the two bonded atoms have different electronegativities.

CPT (*theorem*): general theorem of physics assuming that physical laws are unchanged if, simultaneously, space is reversed (parity operation), time is reversed (the sense of motion is reversed) and matter is replaced by antimatter. In many cases but not all, the CP theorem alone is valid.

Craton: Large block of old (often Precambrian in age) and very stable continental crust (see also shield).

Crust (*geology*): The more superficial shell of the solid Earth. Its lower limit with the mantle is called the Mohorovicic (Moho) discontinuity. Together with the rigid part of the upper mantle it forms the lithosphere. Two main crusts exist: (1) oceanic crust, basaltic in composition and about 7km thick, it constitutes the ocean floor; and (2) continental crust is granitic in composition, with a thickness ranging between 30 and 80km, it forms the continents.

Cryosphere: Part of the Earth's surface made of ice.

Cumulate: Igneous rock generated by accumulation of crystals extracted from magma.

Cyanamid: $\text{NH}_2\text{-CO-CN}$.

Cyanhydric acid: HCN, hydrogen cyanide. Triatomic molecule that during prebiotic period could have played the role of starting material for purine synthesis.

Cyanoacetylene (or better cyanoethyne): H-CC-CN .

Cyanobacteria: Microorganism belonging to the Bacteria domain and able to perform oxygenic photosynthesis. In the past, these microorganisms were improperly called blue algae. The cyanobacteria could be the ancestors of chloroplasts.

Cyanogen: C_2N_2 .

Cysteine (*Cys*): Proteinic amino acid containing three C atoms and a $-\text{SH}$ group in its side chain. In a proteinic chain, two cysteine residues can be linked together by a $-\text{S-S}-$ bond (disulfide bond) often used to stabilize protein conformation. Two cysteines linked together by a disulfide bond is a cystine molecule.

Cytidine: The ribonucleoside of cytosine. The corresponding deoxyribonucleoside is called deoxycytidine.

Cytochrome c: One particular example of the large cytochrome family. Cytochrome c is a protein involved in the electron transfers associated to aerobic respiration. In the eukaryotic cells, cytochrome c is localized in the mitochondrias.

Cytoplasm: Whole content of a cell (protoplasm) except the nucleus whose content is called nucleoplasm.

Cytoplasmic membrane: Also called plasmatic membrane or cell membrane.

Cytosine (*C*): One of the nucleic bases of pyrimidine type.

D

D/H ratio: D = Deuterium; H = Hydrogen. Due to their mass difference, molecules containing either H or D are able to fractionate. For instance, during vaporization, a molecule containing the light isotope (H) is more efficiently vaporized than its heavy equivalent (D). Chemical reactions, including biochemical ones, are able to fractionate these isotopes. Consequently, the D/H ratio can provide information on the biotic or abiotic origin of some organic molecules. The D/H ratio can be very high in chondritic organic matter due to D-enrichment during reactions taking place in the interstellar clouds.

Dalton (*Da*): Molecular mass unit equal to the sum of the atomic masses given in amu (atomic mass unit).

DAP: Abbreviation frequently used to design two different molecules, diamino propionic acid and diamino-pimelic acid.

Darwin: Research program from the European Space Agency devoted to the search of extrasolar planets and the study of their atmosphere composition by spatial interferometry, in the infrared spectral region. Five independent telescopes of 1.5m each constitute the basis of this very sophisticated system.

Daughter molecule (*in comet*): In the cometary coma, any molecule produced by photodissociation of a parent molecule coming from the nucleus is called daughter molecule.

Deamination: Reaction associated to the elimination of an amine group (NH_2 , NHR or NRR').

Decarboxylation: Reaction associated to the elimination of a CO_2 molecule.

Decay constant (λ): Probability that one radionuclide will decay over a given time interval. This probability is unique for each isotope and its value is independent of its chemical and mineralogical environment. The constant (λ) unit is s^{-1} .

Deccan (*Trapps*): Voluminous stacking of basaltic flows emplaced in Northwest India at the end of Cretaceous period. It constitutes the evidence of an extremely important volcanic event contemporaneous with the extinction of dinosaurs.

Delta (*isotopic*) (δ): Difference between the isotopic ratio of a sample (R_e) and that of a standard (R_s). $\delta = 1000 \cdot (R_e - R_s)/R_s$, e.g., $\delta^{18}\text{O}$.

Denaturation: Change of the native conformation of a biopolymer. More specifically and in the case of proteins, denaturation can be induced by increasing temperature and/or pressure or by adding a chemical reagent (like urea). Denaturation can be reversible or irreversible. It is generally associated with a loss of enzymatic properties

Denitrification: Capability of some organisms to transform dissolved nitrate (NO_3^-) into molecular nitrogen (N_2) or nitrogen oxide (N_2O). This metabolism requires anoxic or dysoxic conditions.

Dense clouds: See interstellar clouds.

Dense core (*astronomy*): Gravitationally bound substructure located inside a molecular cloud and surrounded by protoplanetary disks. Dense core collapsing leads to the formation of one or several stars.

Deoxyribose: Ose or monosaccharide having a structure identical to ribose except that the OH group in position 2' is replaced by an H atom. In living organisms, deoxyribose comes from ribose via a reduction reaction.

Depletion: Impoverishment of chemical element abundance when compared to a standard of reference composition.

D-glucose: D enantiomer of glucose. It plays a primeval role in living cells where it produces energy through both anaerobic (fermentation) and aerobic

(associated with respiration) metabolism. D-glucose is also a precursor of several other molecules such as oses and nucleotides.

Diagenesis: Chemical and/or biochemical transformation of sediment after its deposition. This process, which generally consists in cementation and compaction, transforms a running sediment in a compact rock.

Diapirism: Gravity driven magma or rock ascent in the Earth. Generally low density materials rise up into greater density rocks.

Diastereoisomers (or diastereomers): Stereoisomers that are not enantiomers. Diastereoisomers are characterized by physical and chemical properties, which can be as different as observed for isomers having different connectivity (constitutional isomers).

Diazotrophy: Capability of some organisms to assimilate nitrogen as N_2 . This assimilating process is also called nitrogen biofixation.

Differentiation (chemistry): Separation of an initially homogeneous compound into several chemically and physically distinct phases.

Differentiation (Earth): Separation from a homogeneous body of several components whose physical and chemical properties are contrasted. In the case of Earth these components are core, mantle, crusts, hydrosphere and atmosphere.

Diffuse clouds: See interstellar clouds.

Dinitrogen: N_2 , also frequently called nitrogen.

Dioxygen: O_2 , also frequently called oxygen.

Disk (of second generation): Disk around a star resulting from the breaking off of solid bodies previously formed like planetesimals, asteroids or comets. The beta Pictoris disk is probably a secondary disk.

Dismutation: Reaction that from a single reactant gives two products one of them being more oxidized than the reactant and the other one more reduced. A typical example is the transformation of an aldehyde into an alcohol and an acid.

Dissimilative metabolism: In a cell, mechanism that uses a compound in energy metabolism. The reduction reaction releases energy (exothermic).

Dissolve inorganic carbon equilibrium: Concentration proportions of the main chemical forms of the dissolved inorganic carbon in aqueous solution: $HCO_3^-/CO_3^{2-}/CO_2$. These proportions depend on the physicochemical conditions, especially the pH.

Disulfide bond: Covalent bond between two S atoms. When the $-SH$ groups of two cysteines react with each other in presence of an oxidant, it gives a cystine molecule, i.e., a cysteine dimer linked by a $-S-S-$ bond. Disulfide bridges stabilize the ternary and quaternary structures of proteins. Some irreversible denaturing of proteins can be associated to form disulfide bonds.

DNA: Desoxyribonucleic acid, long chain polymer of desoxynucleotides. Support of the genetic code in most living cells. Frequently observed as a double helix made by two complementary strands

Drake (*equation of*): Empirical formula (containing several adjustable parameters) that according to his author, gives a rough estimation of the number of intelligent civilizations in the galaxy.

Dust (*interstellar, cometary*): Small solid particles (0.1–1 mm), generally made of silicates, metal ions and/or carbonaceous matter.

Dysoxic: Environment (water, sediment, etc.) where oxygen is in limited amount. Synonym: dysaerobic.

E

e (*astronomy*): Orbital parameter that measures the eccentricity of elliptical orbit.

Eccentricity: Parameter (e) characterizing the shape of an orbit. e is equal to 0 for a circle, equal to 1 for a parabola, higher than 1 for a hyperbola and between 0 and 1 for an ellipse.

Ecliptic: Geometric plane of the Earth orbit. More precisely, average planar of Earth-Moon barycenter orbit.

Eclogite: High degree metamorphosed basalt. It consists in an anhydrous rock made up of pyroxene and garnet.

Ecosystem: Community of organisms and their natural environment: Ecosystem = community + biotope.

Eddington: ESA project devoted to asteroseismology and search for extrasolar planets by the transits method. Kepler: similar project for exoplanetary science.

Electrophoresis: Analytical method used in chemistry and based on the difference between diffusion rates of ions when placed in an electric field. Initially the ions are adsorbed on a support or immersed into a viscous medium. Capillary electrophoresis is a technique adapted to the analysis of small samples.

Enantiomeric excess (*i.e., in percents*): In the case of a mixture of two enantiomers whose respective concentrations are D and L with D greater than L , *i.e.*, $= (D - L / D + L) \times 100$.

Enantiomeric ratio: In the case of a mixture of two enantiomers, the enantiomeric ratio is the ratio of the two enantiomers.

Enantiomers: Two stereoisomers that differ only due to their opposite chirality (like an idealized left hand and an idealized right hand).

Enantiomorph: Two objects that differ only due to their opposite chirality (like an idealized left hand and an idealized right hand).

Enantioselectivity: A reaction leading to products that are chiral is said to be enantioselective if it gives an excess of one enantiomer. Enantioselectivity

is generally induced by a chiral reactant or a chiral catalyst (like an enzyme). Enantioselectivity is also used to describe a reaction such that starting from a mixture of enantiomers; one of them reacts faster than the other. In this case, the reactant must be chiral.

Enantiotopic (*chemistry*): A planar molecule like H-CO-CH₃ can be seen as a scalene triangle with summits of different colors. In the 2D space, such an object is chiral and the two faces are said to be enantiotopic. A chiral reagent is able to differentiate enantiotopic faces.

Endergonic: A chemical reaction or a physical change is endergonic if it requires a supply of free energy from its surroundings to succeed.

Endogenous (*biochemistry*): Taking place inside. For instance an endogenous organic synthesis is a synthesis that occurs in a planetary atmosphere or at the bottom of an ocean. Antonym: exogenous.

Endogenous (*geology*): Word that refers to petrogenetic mechanisms taking place inside the Earth, as well as to the rocks generated by such mechanisms. A magmatic or a metamorphic rock is an endogenous rock. Antonym: exogenous.

Endolithic (*biology*): Microorganisms living in rocks.

Endosymbiosis: Process by which an eukaryotic cell lives in symbiosis with an other cell (generally a prokaryotic cell) located in its cytoplasm. Chloroplasts and mitochondria are considered as vestiges of endosymbiotic prokaryotes.

Endothermic: A chemical reaction or a physical change is endothermic if it needs a supply of energy from its surroundings to occur.

Enstatite: Mineral, Inosilicate (simple chain silicate), Mg-rich Pyroxene [MgSiO₃].

Envelope (*circumstellar*): In astronomy, cloud of dust particles and gas surrounding a new-born star (protostar).

Enzyme: In biochemistry, a molecule that catalyzes a reaction. Most enzymes are proteins but some are polynucleotides (ribozymes).

Epitope: Region on a macromolecule that is recognized by an antibody. It is generally about 5 to 12 amino acids long, which is the size of the antigen binding site on the antibody.

Equator (*celestial*): Plane perpendicular to the rotation axis of the Earth and corresponding to an extension in space of the terrestrial equator.

Escape velocity: Velocity required for a body to escape the planetary gravity field. On Earth, the escape velocity is 11km/s.

***Escherichia Coli*:** Bacteria found in the intestine and commonly used in experimental bacteriology. Its size is of the order of one micron. Its genome codes, approximately, for 3000 proteins.

Ester: Molecule that can be described as the result of condensation of an acid and an alcohol associated with elimination of a water molecule. (-O-CO-) is the ester bond.

Eubacteria: Sometimes used instead of bacteria, in order to point out the difference between Bacteria and Archaea. Archaeas themselves are sometimes called Archebacterias.

Eukarya: One of the three domains of life (together with bacteria and archaea). Eukaryotic cells are members of the Eukarya domain.

Eukaryote: Any organism from Eukarya domain and characterized by a nucleus (containing the genetic material) separated from the cytoplasm by a membrane. Eukaryotes can be unicellular or pluricellular.

Europa: Satellite of Jupiter with a diameter of about 3100km (same size as the Moon). An ocean could exist beyond its icy surface.

Evaporite: Sedimentary rock generated by evaporation of huge volumes of water; chemical elements dissolved in water precipitate leading to the deposition of minerals such as halite (NaCl) or gypsum = (CaSO₄, 2H₂O).

Exergonic: Chemical reaction or a physical change is exergonic when it provides free energy to its surroundings.

EXOCAM: Special reactor used to experimentally simulate exobiological processes.

Exogenous (*biochemistry*): Taking place outside. For instance exogenous molecules are molecules synthesized in an extraterrestrial environment and imported on Earth by (micro) meteorites or interplanetary dust particles. Antonym: endogenous.

Exogenous (*geology*): Word that refers to petrogenetic mechanisms taking place at the surface of the Earth (in atmosphere or hydrosphere), as well as to the rocks generated by such mechanisms. A sedimentary rock (sandstone, limestone, etc.) is an exogenous rock. Antonym: endogenous.

Exon: Sequence of transcribed nucleotides that is present in natural RNA and that corresponds to a DNA sequence. In DNA, exons are separated by introns (intervening sequences). Exons and introns are respectively coding and non-coding sequences for proteins.

Exoplanet: Planet orbiting around a star other than the Sun. The number of already discovered exoplanets is greater than 100 (more information on this fast developing field is available on www.obspm.fr/planets).

Exoplanetology: Part of astrophysics devoted to extrasolar planets.

Exosphere: The outermost region of a planet's atmosphere. This is the place where most of the photodissociation reactions of molecules take place. From the exosphere, the molecules with sufficient velocity can escape the Earth's gravitation.

Exothermic: A chemical reaction or a physical change is exothermic if it provides energy to its surroundings.

Explosive nucleosynthesis: Nucleosynthetic reactions taking place during star explosion. The characteristic time of these reactions is by far shorter than that of the same reaction occurring in a star at rest.

Extinction (*interstellar, atmospheric*): Decrease of the light intensity of a star due to light diffusion or absorption by a medium (planetary atmosphere, interstellar cloud). In an interstellar cloud, visible magnitude can be reduced by a factor of 1 while the reduction can be as large as 100 in a protostar dense nucleus.

Extremophile: Microorganism that optimally lives (or optimally grows) in extreme physicochemical environments (P, T, Ph, etc.).

F

Fatty acid: Carboxylic acid R-COOH where R is a long chain containing only C and H atoms.

Fayalite: Mineral, Nesosilicate (isolated SiO₄ tetrahedrons), [(Fe)₂SiO₄]. This mineral is the ferrous end-member of olivine series (peridot family).

Feldspar: Mineral family, Tectosilicate (3D silicate), Feldspars are subdivided into two main chemical groups: (1) Alkali feldspars (NaAlSi₃O₈ = Albite and KAlSi₃O₈ = Orthoclase); and (2) Plagioclase feldspars (NaAlSi₃O₈ = Albite and CaAl₂Si₂O₈ = Anorthite). These minerals constitute 52% of the continental and oceanic crusts.

Fermentation: Biochemical process such that complex organic molecules (i.e., glucose) are transformed into low molecular mass molecules (i.e., ethanol) by cells, in anaerobic conditions. Fermentation corresponds to an oxidation process but the final electron acceptor is an organic molecule instead of oxygen. During fermentation and respiration, ATP is produced but less efficiently.

Ferrihydrite: Iron hydroxide, 5Fe₂O₃ 9H₂O.

Fisher: German chemist who was the first to introduce the D/L nomenclature to differentiate enantiomers and to characterize their absolute configurations.

Fisher–Tropsch (*reaction of*): Reaction that produces hydrocarbons from a mixture of H₂ and CO. The FT reaction had and still has a great industrial importance but it could also have been important in prebiotic chemistry. The FT reaction requires metallic catalysts.

Flint (or flintstone): See chert. Rock mainly made of amorphous silica and having a biogenic origin. It frequently appears as nodules in chalk or limestone.

Fluid inclusion: One to 100 μm-sized cavities in minerals containing fluids trapped during mineral crystallization.

Fluorophore: Fluorescent molecule. This molecule absorbs light and its electrons are excited to higher energy states; their return to lower energy states is accompanied by light emission (fluorescence).

Formaldehyde (or methanal): The simplest aldehyde (H-CO-H).

Formation (*geology*): Group of terrains or rocks having the same characteristics (specific lithology or group of lithologies).

Formic acid: HCOOH, methanoic acid.

Formose (*reaction*): Starting from formaldehyde in water solution at high pH, this reaction leads to the formation of a large variety of sugars. Its importance in prebiotic chemistry remains an open question. It is also called the Butlerow reaction.

Forsterite: Mineral, Nesosilicate (isolated SiO₄ tetrahedrons), [(Mg)₂SiO₄]. This mineral is the magnesium end-member of olivine series (peridot family).

Fossil (*geology*): All kinds of traces of passed life (bone, shell, cast, biomolecule, track, footprint, etc.).

Fractionation (*chemistry*): Separation of chemical elements or isotopes by physical or chemical mechanisms.

Frazil: Ice disks with a diameter of few mm that are observed in water as soon as surfusion occurs. Frazil is common in Arctic and Antarctic rivers but also below the huge ice platforms in the Antarctic Ocean.

Free-fall time: Time required for an object of mass m , initially at rest, to reach an object of mass M ($M > m$) under effect of gravitation alone. It gives a good approximation of the time required for an accretion disk to collapse during the protostar stage (typically 10⁵ years).

FRET (*Fluorescence Resonance Energy Transfer*): The energy of an excited electron in one molecule (the donor) can be passed on to an electron of another molecule in close proximity (the acceptor) through resonance, then released as a photon. If the donor is on a target and the acceptor on a ligand, a molecular recognition event can be detected by the photon emission resulting from the ligand-target complex formation.

Furanose: A furanose ring is a cyclic ose formed of four carbons and an oxygen atom.

G

Ga: Gigaannum = one billion years (Gy).

Gabbro: Plutonic magmatic rock. It has a granular texture and mainly consists in pyroxenes and plagioclase (\pm olivine). Basalt is its effusive equivalent.

Gaia: Ambitious project of ESA to measure the position of one billion of stars with a precision of one micro arc second. Gaia is essentially devoted to the search for extraterrestrial planets.

Galactic open cluster (*astronomy*): Cluster that can contain from a dozen to a few thousand of stars (I population) and is younger than a globular cluster.

Garnet: Mineral, Nesosilicate (isolated SiO_4 tetrahedrons), Its general composition is $\text{Y}_2^{+++}\text{X}_3^{++}(\text{SiO}_4)_3$ ($\text{Y} = \text{Al}, \text{Fe}^{+++}, \text{Cr}$ and $\text{X} = \text{Ca}, \text{Mg}, \text{Fe}^{++}, \text{Mn}$). Garnet is stable at high pressure ($>70\text{km}$ depth). In the mantle, garnet is the only aluminum-bearing phase.

GC (*Gas Chromatography*): Chromatographic method using gas as moving phase. It can be used in analytical and preparative chemistry.

Gene: Segment of DNA, containing hundreds to thousands of nucleotides, found in a chromosome. A gene codes for a specific protein.

Genome: Ensemble of genes of an organism.

Genomic: Science that studies genomes. Genomic includes studying and sequencing of genomes as well as method for the analyses of mRNA (transcriptome) and proteins (proteome). Genomic also tries to index the genes of an organism, to localize them on chromosomes, to determine their sequence and to study their functionality.

Genotype: Ensemble of the genetic characters of an organism.

Geocruiser: Asteroid with an orbit that intersects that of Earth. Objects on such an orbit may eventually collide with Earth.

Geographic pole: Point where the rotation axis (instantaneous) of a planet intersects the surface.

Geomorphology: Branch of the geosciences that studies the characteristics, the configuration and the evolution of land forms. The methods developed on Earth to interpret the geomorphologic mechanisms (magmatism, erosion, sedimentation, etc.) are especially well-adapted to the interpretation of images from other planets, where remote sensing is the only available source of information.

Geothermal gradient: Thermal gradient corresponding to the temperature increase with depth. In the Earth crust the geothermal gradient is $\sim 30^\circ\text{C km}^{-1}$.

Giant planets: Large size planets of low density, such as Jupiter, Saturn, Uranus and Neptune. One can distinguish two groups: gaseous giants, Jupiter and Saturn, mainly made up of gas (H_2 , He) coming from the protosolar nebula and icy giants, Uranus and Neptune, rich in ices (H_2O , NH_3 , and CH_4). They all have a core made of heavy elements and were formed in the outer part of the solar nebula, beyond the ice line.

Glacial-interglacial alternation: Oscillatory change of ice volume, mainly due to the extension of ice caps located in high latitudes. The episodes of large ice cap extension are called glacial, while those of minimum extension are called interglacial. These oscillations are not a permanent feature of the Earth's climate; they require specific conditions, in term of plate tectonic configuration (position of the continents), and orbital parameters. The present-day climate is interglacial.

Glaciation: Cold period in the Earth history characterized by the presence of a large cryosphere (ice). Frozen water accumulates and forms ice caps on the continents; consequently, it becomes unable to return to ocean whose level decreases. The main glaciations occurred during the Precambrian, the Early Cambrian, the end of the Ordovician, the Carboniferous, and from the Oligocene to the Quaternary.

Glass inclusion: One to 100 μm -sized cavities in minerals containing magma trapped during mineral crystallization.

Glass: Amorphous material. In volcanic rocks, it can result of the rapid cooling of the magma.

Global ecosystem (planetary ecosystem): The whole set of interactions between the living and non-living compartments of a life-bearing planet, which contribute to regulate its state far from the thermodynamic equilibrium (where no life is possible).

Globular cluster (*astronomy*): Large spherical cluster containing from a few thousand to several million old stars (II population)

Glucide: Name for sugar molecule, also called carbon hydrate, saccharide or ose. In living cells, they can be energy providers and components of nucleic acids or of wall proteins (glycoproteins).

Glucose: $\text{CH}_2\text{OH}(\text{CHOH})_4\text{COH}$. D-glucose is the wider spread ose. Glucose is an aldohexose: the carbonyl group belongs to the aldehyde function (aldose); it contains six atoms of carbon (hexose). L-Glucose is also called levulose and D-Glucose is called dextrose.

Glutamic acid (*Glu*): Alpha AA with a side chain containing an acidic COOH function. Described as a hydrophilic AA.

Glutamine (*Glu*): Amino acid containing five C atoms with a NH_2 group in the side chain, it is considered as hydrophilic.

Glycan: Synonymous to polysaccharides.

Glyceraldehyde: $(\text{HO}-\text{CH}_2-\text{CHOH}-\text{CHO})$; the simplest aldose, containing only one chirotopic carbon atom. The two enantiomers play an historical role in stereochemistry because they are at the origin of the D, L nomenclature, which describes absolute configuration (Fisher). Glyceraldehyde is the biochemical precursor of other oses.

Glycerol: $(\text{HO}-\text{CH}_2-\text{CHOH}-\text{CH}_2\text{OH})$; 1,2,3-propanetriol also called glycerine, component of many membrane phospholipids (which are esters of glycerol).

Glycine (*Gly*): The simplest amino acid and the only one that is achiral.

Glycolic acid: $\text{HO}-\text{CH}_2-\text{COOH}$.

Gneiss: Metamorphic rock made up of quartz, feldspars and micas. All mica crystals show the same orientation thus defining a surface of preferential weakness called a foliation plane.

Gondwana (*Gondwanaland*): Palaeozoic supercontinent formed by convergence and agglomeration of continents (Peninsular India, Madagascar, Africa, Australia, South America and Antarctica) due to plate tectonic activity. It was mainly located in the southern hemisphere.

Gram +: Bacteria, previously colored during a Gram test and that not loses the color after a treatment with ethanol (positive response).

Granite: Plutonic magmatic rock. It has a granular texture and mainly consists of quartz, alkali feldspar and plagioclase feldspar; mica can be present whereas amphibole is rare. Rhyolite is its effusive equivalent.

Granitoid: Family of quartz-bearing plutonic magmatic rocks including granites, granodiorites, tonalites and trondhjemites.

Granodiorite: Plutonic magmatic rock. It is similar to granite but contains not more than 10% of alkali feldspar.

Green bacteria: Microorganisms of the Bacteria domain able to perform anoxygenic photosynthesis.

Greenhouse effect: Warming of a planet surface due to the trapping by planet atmosphere of the electromagnetic waves received and radiated by the planet.

Greenstone belts: Volcanic (basalts and komatiites) and volcano-sedimentary formation widespread in Archaean terrains. It generally presents an elongated shape (~ 100 km long and few tens of kilometers wide). Its green color is due to metamorphism of basalts and komatiites.

Guanine (*G*): Nucleic base with purine structure.

Gy: Gigayear = one billion years (Ga).

H

Habitable zone (*HZ*): Zone around a star where the physical conditions (temperature, presence of liquid water, etc.) on the orbiting planets are considered favorable for the birth and the development of an Earth-like life.

Hadean (*Eon*): Period of time (eon) ranging from 4.55 Ga (Earth formation) to 4.0 Ga (oldest known rock: Acasta gneisses). The Hadean eon belongs to the Precambrian.

Hadley cell: On Earth tropospheric convection cell; the air masses rise at the level of the equator and descend at about 30° latitude. At intermediate latitudes, this cell is relayed by Ferrel cell and by Polar cell at high latitudes. Sometimes, the term Hadley circulation is used to describe the whole system. This global tropospheric circulation results in a redistribution of heat from the equator towards the poles. The concept remains true on Mars as on many other planets. In the case of Mars and Earth, the planet rotation strongly affects the circulation. For instance, the air moving toward the equator in the lower atmosphere

is deflected by the Coriolis effect to create the easterlies trade winds in the tropics.

Half-life (T): Synonym of period. For a single radioactive decay process the half-life is the time needed for the number of radioactive atoms to decrease to half its original value. Half-life (T) is linked to the decay constant (λ) by the relation $T = \ln(2)/\lambda$. T measured in seconds.

Halogen (*chemistry*): Fluorine (F), chlorine (Cl), bromine (Br), iodine (I), astatine (At). Chemical element belonging to the (VIIA) period in the Mendeleev periodic table.

Halophile: Microorganism that lives optimally (or can only grow) in environments with high salt concentration (1M NaCl).

Hamiltonian (operator): Mathematical entity describing the motion of a classic or quantic particle. In quantum mechanics, this mathematical entity allows to write the wave equation for a stationary state whose solutions give the probability density of one or several electrons in any point in space.

Hapten: Can be considered as an isolated epitope. By definition a hapten is a small molecule (few hundred Da), not antigenic by itself: alone hapten cannot induce immune response; it stimulates production of antibodies only in combination with a specific protein called carrier or schlepper.

Harzburgite: Peridotite made up of olivine and orthopyroxene. It generally corresponds to residual mantle after lherzolite melting and extraction of basaltic magma.

HD 209458b: First exoplanet, whose detection by radial velocimetry method was confirmed by the observation of a transit in front of its parent star (HD 209458).

Heavy element (*astronomy*): Any element other than hydrogen and helium.

Helium: Rare gas whose ^3He isotope is used in geology as marker of a recent degassing process from the deep mantle.

Hertzsprung–Russel diagram (*HR diagram*): In astronomy, a 2D diagram with star temperature (or spectral type) as abscissa and star luminosity (or absolute magnitude) as coordinates. Temperatures decrease from left to right and the spectral types sequence is OBAFGKM.

Heterocycle: Cyclic organic molecule containing heteroatoms (i.e., atoms other than C.) as constituents of the cyclic structure.

Heterosphere: Part of an atmosphere located above the homopause and where each gas density distribution decreases according to its own scale height.

Heterotrophous: Organism using reduced organic molecules as principal carbon source for its biosynthesis. Nowadays, these reduced organic molecules are generally produced by other organisms. At the beginning of life, these reduced molecules were probably found in the environment.

Histidine (*His*): Proteinic amino acid containing an imidazole group in its side chain. Histidine residues (hystidyl) are frequently found in the active site of enzymes.

HMT: ($C_6H_{12}N_4$); hexamethylenetetramine, could be a minor component of comet nucleus.

Homeostasis (*biology*): In living organisms homeostasis is the regulation of a physical or a chemical factor that attempts to actively keep these factors at equilibrium, even in case of environmental change (i.e., thermal regulation). This involves a feedback mechanism.

Homeostasis: Property of a living organism to maintain unchanged some of its physicochemical characteristics even in presence of a change in the environment. Homeostasis requires autoregulation.

Homoacetogens: Microorganisms of the Bacteria domain producing acetate from H_2 and CO_2 .

Homochirality: Of the same chirality. All proteinic amino acids are L while ribose in RNA or ATP are always D. The origin of homochirality for the large majority of the chiral constituents of organisms remains an active research subject.

Homogenization temperature: Temperature at which a fluid inclusion transforms from a multiphase (heterogeneous, for instance gas + seawater) to a one-phase (homogeneous) state. This temperature is considered as the minimal temperature of formation of the fluid inclusion.

Homology (*biology*): Two structures in two different species are homologous (and therefore comparable) irrespective of their forms, if they are connected in the same way to identical structures.

Homolysis: Homolytic cleavage or homolytic fission. Breaking of a bond such that each molecule fragment retains the same amount of binding electrons; consequently, no electric charge is created due to the cleavage.

Homolytic (rupture): Rupture of a chemical bond in which the bonding electrons are equally partitioned between the two fragments; no charge appears during this process.

Homopause: Atmospheric boundary between the homosphere and the heterosphere.

Homosphere: Part of an atmosphere where the gases are uniformly mixed. Their densities decrease as the altitude increases following a single mean scale height, defined by the mean molecular weight and the temperature.

Hornblende: Mineral, Inosilicate (double-chain silicate), $[Na_{0-1}Ca_2(Fe^{++},Mg)_{3-5}(Al,Fe^{+++})_{0-2}Si_{8-6}Al_{0-2}O_{22}(OH,F)_2]$. It belongs to the amphibole group.

Hot Jupiter: Massive Jupiter-like exoplanet, orbiting close to a star. Most of the extrasolar planets so far discovered belong to this type.

Hot spot: See mantle plume.

HPLC (*High Performance Liquid Chromatography*): Very efficient liquid phase chromatography performed under high pressure up to 100–400 bars.

Hydrocarbons: Molecules containing only C and H atoms. If the hydrocarbon contains an aromatic system, it is said aromatic. If the hydrocarbon contains only tetra-coordinated C atoms, the hydrocarbon is called aliphatic. Some hydrocarbons result from the polymerization of isoprene (2-methylbutene); they are called isoprenoid hydrocarbons. Latex contains isoprenoid hydrocarbons.

Hydrogen bond: Intermolecular but sometimes intramolecular low energy bond (about 20kJ/mol) involving generally an H atom linked to an electronegative atom like O, N, S and an atom bearing non-bonding electron pairs such as O or N. The H-bond implies an H-donor and an H-acceptor.

Hydrogen cyanide: H-CN.

Hydrogen sulphide: H₂S.

Hydrogenoid (function): Exact electronic wave function describing the motion of the electron in an atom with a single electron.

Hydrolysis: Cleavage of a molecule due to reaction with H₂O.

Hydrosphere: The whole water available on Earth's surface; it includes, oceans, seas, lakes rivers, underground waters and atmospheric water vapor.

Hydrothermal vent: See black smoker.

Hydroxy acid: Carboxylic acid containing also an alcohol function. Glycolic acid is the simplest hydroxy acid.

Hydroxyl (*group*): -OH or alcohol group.

Hyperthermophile: Microorganism living optimally (or only growing) in high temperature environments (>80°C).

Hypoxanthine (*6-hydroxypurine*): Purine base and biological precursor of adenine and guanine.

I

Ice shelves: Ice platforms, generated by glaciers, whose ice progresses extending over the ocean. The Ross ice shelf is about 1000km wide and extends over 600km in ocean. They are the source of very large tabular icebergs.

Ices (*astronomy*): Solid form (crystalline or amorphous) of volatile molecules like water, carbon dioxide or ammonia.

IDP: Interplanetary dust particle.

Igneous rocks: Magmatic rocks due to magma crystallization.

Immunoaffinity chromatography (IAC): All chromatographic methods based on a stationary phase that contains antibodies or fragments of antibodies.

Immunogenic: Substance able to induce an immune response.

Impact melt rock: Rock associated with impact craters (impactite). It resembles a lava and consists of a mixture of solid or molten fragments from the local lithologies floating in a glassy microcrystalline or re-crystallized matrix. It is generated by cooling of the magma produced by the melting of the impacted lithologies under the high temperatures and pressures induced by the collision.

Impactite: Lithologies generated during a meteorite impact. Melt rock, suevites and breccias found inside or at close proximity of an impact crater are called impactites.

Inclination (1) (*astronomy*): Angle between the orbital plane of a solar object and the ecliptic plane in degrees (always lower than 90°).

Inclination (2) (*astronomy*): Angle between the orbital plane of an interstellar object and the sky plane (plane perpendicular to the line of sight, i.e., the straight line joining the observer and the stellar object).

Indels: Acronym for insertions/deletions. Phylogenetic analysis on several DNA or protein sequences requires sequences with the same length. During the process of alignment, alignment gaps (indels) must be introduced in sequences that have undergone deletions or insertions.

Interferometry: Observation and astrophysical measurement technique based on the use of several disconnected telescopes spread over an area, and reaching of the angular resolution that a single-dish telescope with the same area. This technique can be implemented in different ways: visibility measurement, direct imaging by aperture synthesis, differential phase measurement, nulling interferometry, etc.

Interplanetary dust: Small grains left behind by asteroids and comets, and dispersed in a cloud including the whole solar system.

Interstellar cloud: Cloud of gas (98%) and dust (2%). The gas is mainly H (diffuse cloud) or H_2 (molecular cloud). Molecular clouds are called dense clouds ($n(H_2) > 10^3$ molecules cm^{-3}) or dark clouds, if dense and cold (10–20K).

Intertidal environment: Environment between high and low tide. Also called tide range.

Intron: Non-coding sequence of nucleotides that separates exons. Introns are removed during the maturation processes of the three types of RNA by splicing.

Ion-molecule (*reaction*): Kind of reaction between two gaseous reactants and initiated by ionizing cosmic rays, X-rays or UV radiation. They are important in interstellar clouds and in planetary ionospheres.

IR: Infrared. Wavelength ranging between 1 and $300\mu m$.

Iridium (*Ir*): Element belonging to platinum element family. Its concentration in the Earth's crust is extremely low. A local Ir enrichment as at the Cretaceous-Tertiary (K/T) boundary is interpreted as a strong argument in favor of meteoritic impact.

ISM: Interstellar medium.

Isochron (geology): Rectangular diagram plotting isotopic ratio of a disintegration system (abscissa = parental isotope; ordinate = daughter isotope) (e.g., $^{87}\text{Sr}/^{86}\text{Sr}$ versus $^{87}\text{Rb}/^{86}\text{Sr}$). In this diagram, cogenetic rocks of the same age plot along a straight line whose slope is proportional to the age. This method of age determination is widely used in geology.

Isocyanhydric acid: $\text{H-N}=\text{C}$.

Isocyanic acid: $\text{HN}=\text{C}=\text{O}$.

Isoleucine (*Ile*): Proteinic amino acid containing six carbon atoms and described as hydrophobic. Ile is considered as one of the prebiotic AA.

Isoprenoid hydrocarbon: Hydrocarbon formed by polymerization of isoprene $\text{eCH}_2=\text{C}(\text{CH}_3)-\text{CH}=\text{CH}_2$.

Isostasy (geology): Hydrostatic equilibrium in the Earth's crust such that the forces tending to elevate landmasses balance the forces tending to depress them. When this equilibrium is broken, the return to equilibrium takes place by vertical movements of the crust. For instance, in Scandinavia during the last ice age the weight of the ice caps pushed the continental crust into the mantle. As the ice melted, the load decreased and the lithosphere bounced back towards equilibrium level, (this process is still going on at a rate of about 1 m per century).

Isotopic ratio: Concentration ratio of two isotopes or concentration ratio of two isotopomers of a molecule (like H_2O and D_2O). Isotopic ratio can provide information on the age of a sample (when used in isochron calculation) as well as on its origin and source.

Isovaline (*Iva*): Hydrophobic non-proteinic amino acid, isomer of valine and containing five carbon atoms.

Isovaline: Hydrophobic non-proteinaceous amino acid. This constitutional isomer of valine contains five C atoms.

Isua: Region of Greenland where are known the oldest exposed sediments so far recognized (3.865 Ga old gneisses). They contain carbon of possible biogenic origin (see also Akilia).

J

Jeans escape: Process leading to the escape of atomic or molecular species from a planet atmosphere. It happens when the thermal agitation rate is greater than escape rate. The lighter elements or molecules (like H, H_2 or He) escape faster than the heavier ones.

Jovian planets: Other name for giant planets.

Jupiter: The fifth and largest planet (1400 times the Earth volume, 320 times the Earth mass) of the solar system. Jupiter is 5.2AU away from the Sun. Its gaseous envelope, mainly made of H₂ and He, surrounds a core of ice and rocks (10–20 Earth mass).

Juvenile gases: Gases produced by or trapped inside the Earth and reaching the Earth's surface for the first time. ³He is an example of juvenile gas detected in submarine geothermal fluids.

K

K/T (Strata): Few centimeter-thick sedimentary layer located at the Cretaceous Caenozoic boundary. Its enrichment in Iridium (Ir) is interpreted as due to a huge meteoritic impact.

Kepler: Spacecraft NASA mission devoted to detection of Earth-type exoplanets (equipped with a 1-m telescope).

Keplerian rotation: Orbital motion that follows Kepler's laws.

Kerogen: Insoluble organic matter found in terrestrial sediments and in some types of meteorites like carbonaceous chondrites.

Kilo base or kilo base pair (kb): Unit used to measure the number of nucleotides in a gene or a genome: 1000 base pairs of DNA or 1000 bases of RNA.

Komatiite: Ultramafic high-Mg lava. It contains olivine and pyroxene; minerals, which sometimes can have needle or dendritic shapes (spinifex texture). Komatiites were abundant before 2.5Ga and extremely rare after.

Kuiper Belt or Edgeworth–Kuiper: A large ring-shaped reservoir of comets beyond Neptune at about 30 astronomical units (AU) from the Sun.

L

L/D (ratio): For a chiral molecule, the ratio between the L and the D enantiomer concentrations.

Lactic acid: HO-CH(CH₃)-COOH.

Lagrange points: The five points determining the equilibrium position of a body of negligible mass in the plane of revolution of two bodies (e.g., star-planet couple) in orbit around their common gravity center.

Late heavy bombardment: Heavy bombardment of the Moon (and certainly also of the Earth and others telluric planets) that took place between 4.0 and 3.8Ga ago. It could correspond to either the end of a long period of bombardment by asteroids and comets or to a short cataclysmic phenomenon.

Laurasia: Supercontinent formed by convergence and agglomeration of continents (Europe, North America and Asia) due to plate tectonic activity. It was

mainly located in the northern hemisphere. This continent was the result of the breaking on Pangea in two parts at the end of Palaeozoic.

Lava: Magma emplaced as a flow at the surface of the Earth or any other planet.

Leaching: Dragging of soluble elements or particles of a soil by infiltrated water.

Leucine (*Leu*): Proteinic amino acid containing six C atoms and considered as one of the prebiotic AA.

Lherzolite: Peridotite made up of olivine and pyroxenes (ortho- and clinopyroxenes) as well as of an Al-bearing mineral. Its melting generates basaltic magmas leaving a harzburgite residue.

Ligand: Any atom or group of atoms bonded to the atom of interest. As an example, the four ligands of a chirotopic (asymmetric) carbon atom are necessarily different.

Ligase: Class of enzymes that catalyzes the binding between two molecules (or two DNA fragments).

Light-year: Measure of distance used in astronomy, it corresponds to the distance that light runs in one year (0.946×10^{16} m).

Liquidus: Line that, in a composition versus temperature or pressure versus temperature diagrams, separates the domain where crystals and liquid coexist from the field where only liquid exists.

Lithophile (*geochemistry*): Chemical element frequently associated with oxygen. A lithophile element has a greater free energy of oxidation per gram of oxygen than iron; it occurs as an oxide generally in silicate minerals (i.e., Si, Al, Na, K, Ca, etc. (synonym: oxyphile).

Lithosphere: External rigid shell of the Earth. Its definition is based on rheological behavior of rocks. It includes crusts (continental and oceanic) as well as the upper rigid part of the mantle its thickness varies between 0 and 250 km and more or less corresponds to the 800°C isotherm.

Lithotroph: See chemolithotroph. Living organism taking its energy from the oxidation of inorganic molecules such as NH_3 , H_2S , Fe^{++} .

Low-mass star: Star with mass $< 2M_{\text{sun}}$.

L_{sun} (*astronomy*): Sun luminosity (3.826×10^{24} W).

LUCA (*Last Universal Common Ancestor*): Hypothetical microorganism that stood at the root of all evolution lines leading to the present day living beings. Appeared after a long evolution; it cannot be considered as a primitive form of life.

Lysine (*Lys*): Proteinic amino acid containing six C atoms with an amino group in its side chain and which, therefore, is basic and hydrophilic. Lysyl residues are frequently found in the active site of enzymes.

M

Ma: Megaannum = Megayear = one million years (My).

Macronutrient: Chemical element necessary in large amounts for the proper growth and metabolism of a living organism (i.e., C, O, H, N, P, K, S, Ca, Mg, etc.). In seawater, the main macronutrients are water, CO₂ nitrates and phosphates (sources for H, O, C, N and P, respectively). See also micronutrient.

Mag: Magnitude.

Magma: Molten rock that can be completely liquid or consist in a mixture of liquid and crystals. It is produced by high temperature (> 650°C for granite; > 1200°C for basalt) melting of pre-existing rocks. Mantle melting generates basalts, oceanic crust fusion rather generates adakites or TTG and melting of the continental crust gives rise to granites.

Magnesiowurstitute: Mineral, Oxide. [(Mg,Fe)O]. Magnesiowurstitute together with perovskite is probably the main component of the lower terrestrial mantle (depth > 660km).

Magnetic anomaly: Difference between the measured and the theoretical value of the magnetic field intensity of Earth.

Magnetic pole: Point where the magnetic dipole axis of a planet intersects the globe surface.

Magnetite: Mineral, Iron oxide [Fe⁺⁺Fe₂⁺⁺⁺O₄]. Its ferromagnetic properties make it able to record past Earth magnetic field characteristics. It can also exist in some bacteria called magnetotactic.

Magnitude (Mag) (astronomy): Measure of brightness of a stellar object on a logarithmic scale. The difference between two successive magnitudes is a factor 2.512. $\text{Mag} = -2.5 \log_{10} (I/I_0)$. The less bright a star is, the greater its magnitude. The magnitude is calculated on a chosen spectral interval (visible, IR) or on the total spectrum (bolometric magnitude).

Major half-axis: For an elliptic orbit, half of the distance aphelion-perihelion.

Mantle plume (hot spot): Ascending column of hot mantle assumed to be generated near the mantle-core boundary or at the upper-lower mantle boundary. Near surface, this column can melt giving rise to oceanic island magmatism (i.e., Hawaii, La Réunion, etc.).

Mars Express: ESA space mission to Mars.

Mantle: In a planet, the mantle is the shell comprised between crust and core. On Earth, it represents 82% of its volume and 2/3 of its mass, it is divided into upper mantle (until ~700km depth) and lower mantle (until 2900km depth).

Mass loss rate (astronomy): Mass ejected per time unit by a star during its formation. Ejection takes place through stellar winds and bipolar jets (typically 10⁻⁵ to 10⁻⁸ solar mass per year).

Massive star: Star with mass $> 2 M_{\text{sun}}$.

Matrix (*chemistry*): Parent molecule that allows the pre-positioning of isolated elements thus making possible their polymerization.

Maturation (*genetics*): Transformation step of mRNAs leading to their functional form. It occurs by splicing intron sequences.

Megaton: Unit of energy equivalent to the energy released by 10^9 kg of TNT (trinitrotoluene). Corresponds to 4.2×10^{15} J.

Mercaptans: Other name of thiols: sulfur analogues of alcohols.

Mesopause: Atmospheric boundary between mesosphere and thermosphere.

Mesophase: State of matter exhibiting the characteristics of two phases. Liquid crystals have the fluidity of liquids but are characterized by an order at short range similar to what is observed in crystals.

Mesosphere: Atmospheric layer located above the stratosphere and below the thermosphere, between 45–50 km and 80 km (in the case of the Earth) or, if the stratosphere is absent, directly above the tropopause (as in the case of Mars and Venus).

Mesotartaric acid: $\text{HOOC-CHOH-CHOH-COOH}$, alpha-beta-dihydroxysuccinic acid, molecule containing two asymmetric carbon of opposite chirality. This molecule is achiral by internal compensation.

Mesozoic (*Era*): Period of time (era) ranging from 250 Ma to 65 Ma, it was also called the Secondary Era.

Metabolism: All the reactions taking place in a cell or in an organism. Metabolism is divided into two subclasses: anabolism and catabolism. The very large majority of metabolic reactions are catalyzed by proteinic enzymes.

Metallicity: In a star, a galaxy or a gas cloud, the metallicity is the proportion of heavy elements (heavier than helium).

Metamorphism: Solid state transformation of a rock due to change in pressure and/or temperature conditions. New mineral assemblage, stable under the new P-T conditions will appear. New minerals crystallize perpendicular to the oriented pressure thus defining a new planar structure called foliation. Most often metamorphism corresponds to dehydration of the rock.

Metasediments: Sediments affected by metamorphism.

Metasomatism: Change in rock composition due to fluid circulation. For instance, in a subduction zone, the fluids (mainly water) released by dehydration of the subducted slab, up-rise through the mantle wedge. These fluids, which also contain dissolved elements, not only rehydrate the mantle peridotite, but also modify its composition.

Meteor Crater: Impact crater in Arizona. It is about 1.2 km in diameter and 170 m deep. The impact, which took place 50 000 years ago was due to an iron

meteorite of about 25m in diameter. Impacts of this kind generally occur every ~25 000 year.

Meteorite: Extraterrestrial object, fragment of an asteroid, of a planet (like Mars) or of the Moon that falls on the Earth's surface.

Methanogen: Archeobacteria producing methane CH₄ from CO₂ and H₂. Some methanogens are hyperthermophilic.

Methanogen: Methane-producing microorganism of the Archaea domain.

Methionine (*Met*): Proteinic amino acid containing five carbon atoms with a -SCH₃ group in its side chain.

Methylalanine: Synonymous of alpha aminoisobutyric acid (alpha-AIB).

MGS (*Mars Global Surveyor*): American probe that have carried out a complete cartography of Mars (from September 1997).

MHD: Magnetohydrodynamic.

Mica: Mineral family, Phyllosilicate (water-bearing sheet silicate): biotite = black mica [K(Fe,Mg)₃Si₃AlO₁₀ (OH)₂]; muscovite = white mica [KAl₂Si₃AlO₁₀ (OH)₂].

Microarray: Microarrays are small, solid supports on which thousands of different substances (antibodies, proteins, DNA) can be implanted in specific locations. The supports are usually glass microscope slides, but can also be silicon chips or nylon membranes. The substance is printed, spotted, or actually synthesized directly onto the support. Microarrays are used to test large number of targets quickly or when the only small quantities of the sample to be studied are available.

Micrometeorite: Very small meteorite (< 1mm). The 50–400µm fraction is the most abundant found on Earth. Micrometeorites constitute more than 99% of the extraterrestrial material able to reach the Earth's surface (major impacts excepted).

Micronutrient: Chemical element required in small amounts for the proper growth and metabolism of a living organism (i.e., B, Cu, Co, Fe, Mn, Mo, Zn, etc.). See also macronutrient.

Microorganism: Organism invisible without a microscope. Includes prokaryotes and unicellular eukaryotes (i.e., yeasts).

Microspheres: Spherical clusters of organic molecules found in Precambrian rocks or produced in laboratory from amino acids polymers (proteinoids). Today, the Fox proteinoids microspheres are not considered anymore as plausible models of primitive cells.

Microsporidia: Parasitic unicellular eukaryotes that have been shown to be related to fungi from the fungi. Microsporidia were thought for some time to be primitive.

Migmatite: High-temperature metamorphic rock affected by partial melting.

Milankovitch (*theory of*): Theory connecting the Earth climate variations to astronomic variations such as changes of Earth's orbit or obliquity with time.

Miller–Urey (*experiment of*): One of the first experimental simulations of what was considered as atmospheric prebiotic chemistry (1953). Synthesis of a large variety of organic molecules including few amino acids from a very simple mixture containing reduced small molecules (H_2 , CH_4 , NH_3 and H_2O) submitted to an electric discharge.

Mineral: Solid material defined by both its chemical composition and crystalline structure.

Minimal protosolar nebula: Minimal mass of gas, necessary for the formation of the solar system planets (mass of all planets + H + He \approx 0.04 solar mass).

Minor planet: Asteroid or planetoid.

Mitochondria: Organelles in the cytoplasm of all eukaryotic cells where ATP synthesis takes place during aerobic respiration. Mitochondria have their own DNA and could have an endosymbiotic origin.

Mitosis: Nucleus division, cell division step including cytokinesis.

MM: Micrometeorites.

Moho (*Mohorovicic*): Discontinuity in seismic waves that marks the crust-mantle boundary.

Mole: SI unit for amount of substance; it is defined as the number of atoms in exactly 0.012kg of carbon-12 (1 mole of atoms = 6.02×10^{23} atoms; 1 mole of molecules = 6.02×10^{23} molecules).

Molecular beacon: Molecular beacons are single-stranded oligonucleotide hybridization probes that form a stem-and-loop structure. The loop contains a probe sequence that is complementary to a target sequence, and the stem is formed by the annealing of complementary arm sequences that are located on either side of the probe sequence. A fluorophore is covalently linked to the end of one arm and a quencher is covalently linked to the end of the other arm. Molecular beacons do not fluoresce when they are free in solution. However, when they hybridize to a nucleic acid strand containing a target sequence they undergo a conformational change that enables them to fluoresce brightly.

Molecular clouds: See interstellar clouds.

Molecular flow: See bipolar flow.

Molecular recognition: Chemical term referring to processes in which a specific molecule (ligand) adhere in a highly specific way to another molecule (target), forming a large structure.

Monophyly: Term that describes a taxonomic group sharing a single ancestor and all its descendants (i.e., the mammals).

Monosaccharides: See oses.

Montmorillonite: Mineral, Phyllosilicate (water-bearing sheet silicate). Clay mineral belonging to the smectite group.

MORB (*Mid-Ocean Ridge Basalt*): Basalt generated in mid-oceanic ridge systems where oceanic crust is created. Most of ocean floor has a MORB composition.

m-RNA: Messenger RNA. Obtained by transcription of a DNA segment and able to orient the synthesis of a specific protein in the ribosome.

MS (*Mass Spectrometry*): Analytical method involving a preliminary ionization of atoms or ionization and fragmentation of molecules followed by measurement of atomic or molecular masses. These measurements can rely on the precise study of ions trajectories or time of flight in an electric and/or magnetic field.

MSR (*Mars Sample Return*): NASA-CNES space mission project for the return of Martian soil samples (~1kg) extracted by automatic probes. Launch is expected between 2009 and 2014, and samples return three to five years later.

Murchison: Carbonaceous chondrite (CM) fell in Australia in 1969. Fragments recovered immediately after the fall were (and still are) subjected to many analyses, mainly chemical. More than 500 organic compounds were identified, including amino acids and nucleic bases.

Muscovite: Mineral, Phyllosilicate (water-bearing sheet silicate). $[KAl_2Si_3AlO_{10}(OH)_2]$. It belongs to the mica group and is also called white mica.

Mutation: Any change of the genetic material, transmitted to the descendants.

My: Megayear = one million years (Ma).

Mycoplasma: The simplest and the smallest known microorganisms, they live as parasites in animal or vegetal cells. They have long been considered as possible analogues of the first cells; they are now considered as Gram positive bacteria, which lack their rigid cell wall and have evolved by reduction.

N

Nanobacteria: Hypothetical bacteria with size around a few nanometers, smaller than any currently known bacteria. Their existence is very much debated.

N-carbamoyl amino acids: A molecule showing many similarities with amino acids except that one of the H atoms of the amino group is substituted by the carbamoyl polyatomic group (-CO-NH₂). N-carbamoyl amino acids could have been prebiotic precursors of some amino acids.

Neutral (mutations) (*genetics*): Term coming from the neutral theory of molecular evolution proposed by Kimura. A neutral mutation is a mutation leading to sequences selectively and functionally equivalent. They are termed neutral in regards of evolution.

Neutron star: Remnant of a dead star with an initial mass greater than $\sim 8 M_{\text{Sun}}$. When electron degeneracy pressure becomes too low; electrons penetrate into nucleus and are transformed into neutrons by reacting with protons (neutralization). Then they develop a neutron degeneracy pressure.

NGST (*Next Generation Space Telescope*): NASA project of space telescope (4m), it will succeed the HST (Hubble Space Telescope).

Nitrification: Microbial oxidation, autotrophic or heterotrophic, of ammonium to nitrate.

Nitrile: R-CN where the CN group is the cyano group (cyano as prefix but nitrile as suffix).

Non-reducing atmosphere: Atmosphere of CO_2 , N_2 , H_2O where hydrogen is absent or in low quantity, either in the form of free H_2 or hydrogen-containing compounds, such as methane or ammonia. The atmosphere can also be named oxidized or neutral according to its composition.

Non-sense (*codon*): When a codon (triplet of nucleotides) does not specify an amino acid but corresponds to a termination codon (Term.). In the universal code, these codons are UAA, UAG and UGA.

Normative (*rock composition*): Rock mineralogical composition recalculated from its chemical composition.

Nuclear pores: Complex structures, highly specialized, embedded in the nuclear membrane. They allow the transfer of macromolecule between nucleoplasm and cytoplasm.

Nucleation (*astronomy*): Mechanism of formation of solid bodies by accretion of planetesimals.

Nucleic acid: Long chain polymeric molecule obtained by condensation of nucleotides. DNA and RNAs are nucleic acids.

Nucleic base: Linked to ribose or desoxyribose by a hemi-acetal bond, it gives nucleosides. Nucleic bases are purine bases or pyrimidine bases. Nucleosides together with a phosphate group are the subunits of nucleotides. By condensation, nucleotides give the polynucleotides (including DNA and all RNAs).

Nucleides: Constituents of atom nucleus, i.e., protons and neutrons.

Nucleon: Common name for proton or neutron.

Nucleoplasm: Protoplasm within the nucleus of eukaryotes.

Nucleotide: Molecule made by condensation of a base (purine or pyrimidine), an ose and a phosphate group linked to the ose. Nucleotides are ribonucleotides when the ose is ribose or deoxyribonucleotide when the ose is deoxyribose. DNA is a polydeoxyribonucleotide while RNAs are polyribonucleotides. The symbol of a nucleotide is determined by the base (A for adenine, C for cytosine, G for guanine, T for thymine, U for uracil).

Nucleus (*biology*): Eukaryote cell substructure that contains the chromosomes.

O

Obduction: Mechanism leading to the thrusting of oceanic lithosphere onto continental crust.

Obliquity: Angle between the ecliptic and the celestial equator, actually, 23.3 degrees for the Earth. This angle is $> 90^\circ$ if a planet has a retrograde rotation.

Ocean resurgence: See upwelling.

Oceanic rift: Central depression in a mid ocean ridge, this is the place where oceanic plates are created.

Oligomer: Small polymer, generally containing less than 25 monomeric units.

Oligomerization: Polymerization involving a small number of monomers.

Oligopeptides: Small polypeptide (less than 25 AA residues even if the definition is not so strict).

Olivine: Mineral, Nesosilicate (isolated SiO_4 tetrahedrons), $[(\text{Fe}, \text{Mg})_2\text{SiO}_4]$. This mineral belonging to the peridot family, is silica-poor and is one of the main components of the terrestrial upper mantle. It is also common in meteorites.

OMZ: See oxygen minimum zone.

Oort cloud: Huge spherical collection of comets, orbiting the Sun between 10 000 and 100 000 AU.

Ophiolite: Part of oceanic lithosphere tectonically emplaced (obducted) on a continental margin.

Optical activity: Orientation change of the linearly polarized plane of light after its passage through a chiral medium.

Optical rotatory dispersion: Change of the optical power of a chiral medium with the wavelength of the linearly polarized light.

Orbital migration: Change, in course of time, of the planet-star distance; this hypothesis is proposed in order to explain the presence of massive planets close to their star.

Organic molecule: Until the beginning of the 19th century, organic molecules were molecules extracted from plants or animals. Today, any molecule containing carbon atoms is called organic. Carbonates, CO and CO_2 are borderline cases. Organic molecules generally contain C atoms with an oxidation number lower than 4.

Organometallic molecule: Molecule containing one or more metal atoms.

Organometallic: Organic molecule containing one or more metallic atoms bonded by covalent bonds or by coordination to the organic moiety of the molecules. Metallic salts of organic acids are not considered as organometallic compounds.

Orgueil: Large CI carbonaceous chondrite (very primitive), without chondres, that fell in France in 1864, near Montauban.

Ornithine: $(\text{NH}_2\text{-CH})\text{-COOH}\text{-(CH}_2\text{)}_3\text{-NH}_2$. Non proteinic amino acid, precursor of arginine.

Orogenesis: Mountain chain genesis.

Orographic: Related to relief. In weather science, the presence of a relief can induce changes in atmospheric circulation resulting in weather changes (i.e., orographic rainfalls).

Orthose: Mineral, Tectosilicate (3D silicates), Alkali feldspar KAlSi_3O_8 .

Oses (or saccharides): Large group of molecules of primeval importance in the living world. Some of them are monomers like glucose ($\text{C}_6(\text{H}_2\text{O})_6$) or ribose ($\text{C}_5(\text{H}_2\text{O})_5$). Some of them are dimers like lactose or saccharose; some of them are polymers (polysaccharides) like starch or cellulose.

Outer membrane: membrane surrounding the plasma membrane in Gram negative bacteria (i.e., *Escherichia coli*).

Oxidative phosphorylation: In living cells, biochemical process that results in ATP synthesis; energy is provided by protons transfer across the cell membrane.

Oxygen minimum zone (OMZ): Part of the water column where the oxygen consumption due to respiration exceeds its supply by lateral circulation, giving rise to dysoxic or anaerobic conditions. The OMZ areas occur generally where the intermediate or deep circulation is weak, and/or when biological productivity at the surface is high.

Oxyphile: See lithophile.

P

P4: Certification for laboratories accredited to analyze high-risk infectious agents. Such laboratories must be protected against any risk of contamination by viruses and microorganisms, from inside to outside and from outside to inside. These two conditions must absolutely be realized for extraterrestrial sample analyses.

PAH (Polycyclic Aromatic Hydrocarbons): Organic molecules like naphthalene or anthracene containing several fused aromatic rings.

Paleomagnetic scale: Relative stratigraphic scale based on the successive inversions of the Earth magnetic field through time.

Paleosoil: Fossil soil. These formations are able to have recorded O_2 and CO_2 concentrations in the primitive terrestrial atmosphere.

Paleozoic (Era): Period of time (era) ranging from 540Ma to 250Ma, it is also called the Primary Era.

Pangaea: Supercontinent that existed at the end of Palaeozoic era (-225Ma) and later broke in two parts: Laurasia (N) and Gondwana (S).

Paralogous: Paralogous genes originate in gene duplication events, in contrast to the standard orthologous genes, which originate via speciation events.

Parent body: Asteroid or planet from which an object has been extracted, such as for example a meteorite.

Parent molecule (*in comet*): Molecule present in the comet nucleus.

Parity (*violation of*): Characterizes any physical property that changes when space is inverted or, in other words, which is not the same in the mirror world. Parity violation is observed in several phenomena related to the weak intranuclear interactions like the beta radioactivity. Parity violation is at the origin of the very small energy difference between enantiomers (PEVD for Parity Violation Energy Difference).

Parsec (*separ*): Unit of astronomical length of 3×10^{18} cm (about 200 000 AU or 3.26 light-years); it is based on the distance between the Earth and the point where the stellar parallax is 1 second of arc. The average distance between stars in the Sun vicinity is around 1 parsec.

Pathfinder: American probe that landed on Mars on July 4, 1997, formally named the Mars Environmental Survey (MESUR). It contained a rover Sojourner that explored Ares Vallis during several months. For instance, an Alpha Proton X-ray Spectrometer was used to analyze soil and rock samples to determine their mineralogy.

PCR (*Polymerase Chain Reaction*): Experimental method that, by successive molecular duplications, leads to a dramatic increase of a small initial amount of DNA. This result is obtained with an enzyme called DNA-polymerase isolated from thermophile bacteria.

Peptide: Polymer obtained by condensation of amino acids. In a peptide (or polypeptide), the number of AA residues is generally lower than 60. With a higher number of residues, the polymer generally adopts a well-defined conformation and is called a protein.

Peridot: Mineral family, nesosilicate (isolated SiO_4 tetrahedons), olivine is a peridot.

Peridotite: Rock made up of olivine and pyroxenes (ortho- and clino-pyroxenes) as well as of an Al-bearing mineral (spinel at low pressure and garnet at high pressure. Earth mantle is made up of peridotite.

Perihelia: In the case of an object in elliptical motion around a star, the point that corresponds to the shortest distance with respect to the star.

Permafrost: In arctic regions, permanently frozen soil or subsoil.

Perovskite: Mineral, $[(\text{Mg},\text{Fe})\text{SiO}_3]$. Perovskite together with magnesiowurstone is probably the main component of the lower terrestrial mantle (depth > 660 km).

Petrogenesis: Mechanism(s) of formation of rocks.

PGE or platinum group elements: Transition metal belonging to the Platinum group: ruthenium (Ru), rhodium (Rh), palladium (Pd), osmium (Os), iridium (Ir) and platinum (Pt). They possess similar chemical properties (siderophile). In the Earth's crust their abundance is very low, whereas it can be high in undifferentiated meteorites. On Earth, their high concentration in sediments is used as an indication of a meteorite impact, for example at the KT boundary. The elemental ratio between the PGE can provide information on the nature of the impacted meteorite.

pH: Measure of the acidity of an aqueous solution. $\text{pH} = -\log_{10} (\text{H}^+)$ where (H^+) is the molar concentration of hydroxonium ion in solution. $\text{pH} = 7$ corresponds to the neutrality while a solution with $\text{pH} > 7$ is basic and a solution with $\text{pH} < 7$ is acidic.

Phanerozoic (*Eon*): Period of time (eon) ranging from 0.54Ga to today; it follows the Precambrian and is characterized by the development of metazoa.

Phenetic: Taxonomic system for living beings based on overall or observable similarities (phenotype) rather than on their phylogenetic or evolutionary relationships (genotype).

Phenotype: The observable characteristics of an organism, i.e., the outward, physical manifestation of an organism.

Phenylalanine (*Phe*): Amino acid containing an aromatic phenyl group in its side chain. Phenylalanine contains nine amino acid residues.

Phosphoric acid: H_3PO_4 , dihydrogen phosphate. Molecule that plays an important role in the living world: nucleotides are esters of phosphoric acid.

Photochemistry: Chemistry involving energy supply coming from light (from IR to UV). When electromagnetic frequency is greater (X-rays, beta rays or gamma rays), the term radiochemistry is generally used. Specific processes like photoactivation, photoionisation, photodissociation (including the production of free radicals), and photolysis are various aspects of photochemistry.

Photodissociation: Dissociation of molecules due to the energy provided by electromagnetic radiation.

Photolysis: See photodissociation.

Photosynthesis: Synthesis using photons as energy supply. Photosynthesis can be performed at laboratory or industrial level as a subdomain of photochemistry. In biochemistry, the term photosynthesis describes the different biosynthetic pathways leading to the synthesis of molecules under the influence of light. The photons are absorbed by cell pigments and their energy is converted into chemical energy stored in chemical bonds of complex molecules, the starting material for the synthesis of these complex molecules being small molecules like CO_2 and H_2O . It is important to make a clear distinction between the aerobic photosynthesis (also called oxygenic photosynthesis) and the anaerobic photosynthesis. In the first case, water is the reductive chemical species and O_2 is a by-product

of the reaction. It must be kept in mind that atmospheric dioxygen as well as oxygen atoms of many oxidized molecules on the Earth's surface have a biosynthetic origin. Green plants, algae and cyanobacteria are able to perform aerobic photosynthesis; their pigments are chlorophylls. The anoxygenic photosynthesis is not based on H_2O but on reductive species like H_2S . Finally, some halophile archaeas contain a pigment called bacteriorhodopsin and are able to use light energy supply for ATP synthesis. So, light energy is converted into chemical energy or better, in chemical free energy.

Phototroph: Organism whose energy source is light (photosynthesis).

Phylogenesis: History of the evolution of a group of organisms.

Phylogenetic tree: Schematic representation describing the evolutive relationships between organisms. It gives an image of the evolution pattern.

Phylotype: Environmental sequence, representing an organism. Sequence of a clone obtained from environment and representing an organism.

Phylum (*pl. phyla*): Group of organism evolutionary connected at high taxonomic rank.

Pillow lava: Lava extruded under water (ocean, lake), which produces its typical rounded pillow shape. Pillow lava forms the upper part of oceanic crust.

PIXE (*Proton-Induced X-ray Emission*): Device that allows the detection and identification of many elements (metals) in proteins.

Plagioclase: Mineral, Tectosilicate (3D silicates), Calco-sodic feldspar whose composition ranges between a sodic ($\text{NaAlSi}_3\text{O}_8 = \text{Albite}$) and a calcic ($\text{CaAl}_2\text{Si}_2\text{O}_8 = \text{Anorthite}$) pole. They represent about 40% of the Earth crust minerals.

Planet: Body formed in circumstellar disks by accretion of planetesimals and may be of gas.

Planetary nebula: This expression is not related to planets or to nebula. A planetary nebula is the low density and expanding gaseous envelope of a white dwarf. The last phase of evolution of a star whose mass is $< \sim 8 M_{\text{Sun}}$ consists in an explosion (nova) and in the ejection of a gaseous envelope; what remains of the stars produces the white dwarf itself.

Planetesimals: Small solid bodies ($\sim 1 - 100\text{km}$) formed in the protosolar nebula, probably similar to asteroids and comets. Their collision and accretion built the planets.

Plankton (*plankton*): The whole organisms living in water and drifting along ocean and lake currents. It includes zooplankton (small animals) and phytoplankton (plants). Fishes and sea mammals, able to swim independently of current flow, constitute the nekton. Neston refers to organisms drifting at the air-water interface whereas benthos designates organisms living in or on the aquatic ground.

Plasma membrane (or cell membrane): Semi-permeable membrane that surrounds contemporary cells. It consists of a double layer of amphiphilic molecules (hydrophobic tail with hydrophilic head), mainly phospholipids, with proteins embedded in it. It may also contain some molecules, such as cholesterol or triterpenes, able to rigidify the whole.

Plasmid: Extrachromosomal DNA found in bacteria and yeasts, able to replicate independently.

Plasmon: Electronic cloud of a metal.

Plasts: Organites found in phototrophic Eukaryotes and containing photosynthetic pigments such as chlorophylls.

Plate (*lithospheric*): Piece of the rigid lithosphere that moves over the ductile asthenosphere affected by convection.

Plate tectonic: Theory that describes and explains the rigid lithospheric plate motion.

Plutonic: Magmatic rock, resulting of the slow cooling and crystallization of magma at depth. Its texture is granular (big crystals), e.g., granite.

PMS star: New-born star such that the internal temperature is too low ($< 10^7$ K) to initiate nuclear fusion of hydrogen into helium (i.e., T Tauri stars) it is characterized by a low mass.

PNA (*Peptide Nucleic Acids*): Synthetic analogues of nucleic acids such that the ose-phosphate strand is replaced by a polypeptide backbone to which the bases are linked.

Polarized light: It is important to differentiate two limit cases. A linearly polarized light (or plane polarized light) is an electromagnetic radiation, whatever its frequency is, such that the electric vector and thus also the magnetic vector oscillate in a plane. If this radiation travels through a chiral medium, the plane of polarization is deviated (optical rotation). A circularly polarized light is an electromagnetic radiation such that its electric vector and thus also its magnetic vector describe, in space, a helix around the propagation direction. This helix can be right or left corresponding to the two possible circularly polarized lights of a definite frequency. The linearly polarized light corresponds to the superposition of two circularly polarized lights of the same frequency.

Pole motion: Geographic pole motion at the Earth's surface. This motion has low amplitude, since the pole only moves of a few meters.

Polymerization: Chemical reaction such that molecules called monomers are covalently linked together and form long chain molecules, with or without branching. The polymerization can be the result of an addition of polymers like in polyethylene or polystyrene (industrial polymers) but the polymerization can also be the result of condensation reactions involving at each step, the elimination of a small molecule (generally water). Nylon is an industrial condensation

polymer and most of biological macromolecules such as polypeptides, polynucleotides and polysaccharides are condensation polymers.

Polynucleotide: Polymer resulting from condensation of nucleotides.

Polypeptide: Polymer resulting from condensation of amino acids.

Polysaccharide: Polymer resulting from condensation of oses (in the past, oses were also called monosaccharides).

POM (*polyoxymethylene*): Polymer of formaldehyde.

Population I stars: Stars enriched in heavy elements (O, C, N, etc.). They have been formed from the interstellar gas enriched by the previous generation(s) of stars formed over the billions of years of lifetime of the galaxy. They are predominantly born inside the galactic disk. The Sun belongs to population I stars. Massive, hot stars are necessarily young and are therefore always population I stars.

Population II stars: Stars poor in heavy elements (O, C, N, etc.). They have been formed from low metallicity gas that has not been enriched by successive generations of stars. They are believed to be born in the early ages of the galaxy. They are actually distributed predominantly in the halo of the galaxy thus confirming that they are probable remnants of the infancy of our galaxy.

Poynting–Robertson (effect): Effect of stellar light on a small orbiting particle. This causes the particle to fall slowly towards the star. Small particles (below 1 cm) are more affected because the Poynting–Robertson effect is inversely proportional to particle size.

p-p chain: Series of hydrogen burning reactions that produce helium. These reactions provide the energy of the main sequence stars whose mass is $< 1.1 M_{\text{Sun}}$.

ppb (*part per billion*): Relative concentration in mass = nanogram/gram.

ppbv (*part per billion in volume*): Relative concentration in volume = nanoliter/liter.

ppm (*part per million*): Relative concentration in mass = microgram/gram.

ppmv (*part per million in volume*): Relative concentration in volume = microliter/liter.

Prebiotic chemistry: All chemical reactions which have contributed to the emergence of life.

Precambrian: Group of eons ranging from the formation of the Earth (4.55 Ga) until the beginning of Palaeozoic era (0.54 Ga). It includes: Hadean, Achaean and Proterozoic eons.

Precession (*of equinoxes*): Motion of the Earth's rotational axis with respect to the celestial sphere due to the other bodies of the solar system and more particularly to the Moon. The terrestrial pole describes a circle on the celestial sphere in 26 000 years.

Primary structure (of a protein): Sequence of the amino acid residues in the proteinic chain.

Primitive atmosphere: The Earth's primitive atmosphere refers to the atmosphere present when prebiotic chemistry occurred, free oxygen concentration was very low. This atmosphere was considered as highly reductive; today it is rather assumed to have been made up of carbon dioxide, nitrogen and water vapor (greenhouse).

Primitive Earth: The young Earth from its formation until 2.5 Ga.

Primitive nebula: See protosolar nebula.

Prion: Pathological protein containing no nucleic acid.

p-RNA: Synthetic RNA molecule in which the sugar is a pyranose instead of a furanose.

Prokaryote: Microorganism in which chromosomes are not separated from the cytoplasm by a membrane. Bacteria and Archaea are prokaryotes.

Proline (Pro): Amino acid containing six C atoms with a unique characteristic. Its side chain links together the alpha atom and the N atom to give a five membered ring containing one N atom and four C atoms; the amino group is no longer a $-NH_2$ group but a $-NH-$ group. Proline is frequently observed in protein secondary structures called beta turns.

Proper motion (astronomy): Apparent angular movement of a star on the celestial sphere during a year (perpendicular to the line of sight). Proper motion analysis can lead to detection of planets in orbit around this star.

Proper motion: Apparent angular movement of a star on the celestial sphere. The detailed observation and study of such motion can lead to the detection of planets orbiting the star.

Propionaldehyde: CH_3-CH_2-CHO or propanal.

Protein Scaffold: Protein whose main function is to bring other proteins together in order for them to interact. These proteins usually have many protein binding domains (like WD40 repeats).

Proteins: Long-chain biological polymers obtained by condensation of amino acids. The degree of polymerization (number of residues) ranges between 60 and 4000. Some proteins form aggregates. For example, haemoglobin is a tetramer containing two proteinic chains of one type and two proteins chains of another type; in each chain, a heme molecule containing a ferrous cation is settled in without being covalently bonded. Structural proteins are components of muscles or flagella; most enzymes are proteins and proteins are also carriers of other molecules like dioxygen or carbon dioxide. The activity of proteins is extremely dependent on their molecular conformation.

Proteobacteria: Group of bacteria including *Escherichia coli* and purple bacteria (photosynthetic bacteria). Mitochondria are relics of proteobacteria.

Proteome: The complete collection of proteins encoded by the genome of an organism.

Proteomics: Part of science that studies proteome. Proteomics not only analyzes the complete protein collections but also determines their precise location, modifications, interactions, activities and functions. Consequently, it implies the simultaneous analysis of a huge amount of proteins in a single sample.

Proterozoic (*eon*): Period of time (*eon*) ranging from 2.5 to 0.54Ga. This aeon belongs to the Precambrian and follows Achaean. It is characterized by the apparition of oxygen in atmosphere and of metazoa.

Proton motive force: Free energy difference (measured in volts) associated with proton translocation across a membrane. It depends on the electrical membrane potential and on the pH difference between the two reservoirs separated by the membrane. The proton motive force provides energy required for ATP synthesis.

Protoplanetary disks: Disk around a new-born star where accretion of planets is supposed to take place.

Protosolar nebula: Rotating disk of gas, dust and ice, from which the solar system originated.

Protostar: Several similar definitions exist: (1) new-born star such that half of its luminosity is due to accretion, (2) body involved in an accretion process that will bring it on the main sequence, (3) collapsing interstellar cloud, (4) young object that is not yet optically visible. Protostars are rare because their time life is short (10^4 to 10^5 years): Protostars are the Holy Grail of IR and submillimeter astronomy.

Psychrophilic organism: Organism that lives optimally at a temperature lower than 10°C . Some psychrophilic organisms live at temperatures lower than 0°C .

Pulsar: Small neutron star in very fast rotation (one rotation in less than 1s) emitting a highly focalized radiation, circularly polarized and detected as very regular pulses. Pulsars are remnants of supernovae.

Purine bases: Guanine and adenine are examples of purine bases because their molecular skeleton corresponds to purine. These bases are found in DNA as in RNA.

Purple bacteria: Microorganisms of the bacteria domain able to perform anoxic photosynthesis.

PVED (*Parity Violation Energy Difference*): Energy difference between enantiomers due to parity violation at the level of the weak forces.

Pyranose: A pyranose ring is a cyclic ose formed of five carbons and an oxygen atom. It is the more stable form of oses.

Pyrimidine bases: Thymine, cytosine and uracil are examples of pyrimidic bases because their molecular skeleton corresponds to pyrimidine. Cytosine is found in DNA as in the RNA while thymine is specific of DNA and uracil is specific of RNA.

Pyrite: Mineral. Sulphide [FeS₂].

Pyrolysis: Thermal degradation of a molecule.

Pyroxene: Mineral family. Inosilicate (simple chain silicate). Divided in two families: 1) Orthopyroxene [(Fe,Mg)₂Si₂O₆] (e.g., enstatite) and 2) Clinopyroxenes [(Ca,Fe,Mg)₂ Si₂O₆] (e.g., augite, diopside).

Q

Q: Orbital parameter of a planet orbiting a star; distance between planet aphelion and star.

q: Orbital parameter; distance between planet perihelion and star.

Quartz: Mineral, Tectosilicate (3D silicate). It crystallizes in hexagonal or rhombohedral systems. In magma, it characterizes silica sur-saturation. The rhombohedral crystals are chiral: quartz can therefore exist as D- or L-quartz depending from the helicity of the –O-Si-O-Si- chain. A chiral quartz crystal can induce an enantioselectivity during a chemical reaction between achiral reactants via a catalytic effect.

Quaternary structure (of a protein): In the case of protein forming supramolecular aggregates like haemoglobins, the quaternary structure corresponds to the arrangements in space of the subunits.

Quencher: Molecular entity that deactivates (quenches) a fluorophore.

R

R: Solar radius (0.69×10^6 km) or 1/200AU; approximately 10 times the Jupiter radius and 100 times the Earth radius.

Racemate (crystal): Crystalline form of a chiral substance such that each unit cell of the crystal contains an equal number of molecules of opposite chirality.

Racemic (mixture): Mixture of enantiomers containing an equal number of the two enantiomers. Such a mixture is described as achiral by external compensation.

Racemization: Diminution of the initial enantiomeric excess of a homochiral ensemble of molecules or of a non-racemic mixture of enantiomers. It corresponds to an equilibration reaction: the system evolves spontaneously towards a state characterized by higher entropy and therefore, lower free energy. The highest entropy and lower free energy corresponds to the racemic mixture. Racemization can be a very slow process: this is why enantiomers of amino acids, sugars and many other components of living species can be separated in many cases.

Radial velocity measurement: The line of sight velocity of a star or other celestial object towards or away from an observer. This method allows to determine a star's movements through space and consequently to infer the existence (or not) of an orbiting object.

Radial velocity: Star velocity component parallel to the view line. It causes the frequency shift observed in spectral emission lines (Doppler Effect). Periodical change in radial velocity can be an indirect proof that a planet orbits around the star (reflex motion).

Radiation pressure: Pressure applied by electromagnetic waves on any atom, molecule or particle.

Radioactivity (*long-lived species*): Radioactive elements with long periods (10^9 to 10^{11} years); they are still present in the solar system.

Radioactivity (*short-lived species*): Radioactive elements with short period ($< 10^7$ years). Nowadays they have totally disappeared from the solar system.

Raman (*spectroscopy*): Physical method used for molecular structural analysis. Raman spectroscopy is based on inelastic diffusion of visible or UV light and gives information about the vibration modes of diffusing molecules. Raman spectroscopy must be considered as a complementary method with respect to IR spectroscopy.

Rare earth elements (*REE*): Chemical elements with very similar chemical properties. This family (lanthanides) ranges from lanthanum ($Z = 57$) to lutetium ($Z = 71$). In geochemistry, they are commonly used as tracers of magmatic processes. Indeed, they are poorly sensitive to weathering or metamorphism, but on the opposite they are excellent markers of magmatic processes such as melting or crystallization.

Rare gases (*He, Ne, Kr, Ar, Xe, Rn*): Monatomic gases corresponding to the (VIII A) period of Mendeleev periodic table (see Astrobiological data). Their isotopes can be used to trace some geological events of Earth differentiation (e.g., atmosphere and ocean formation).

Recovery ratio (%): A sample prepared by adding a known mass of target analyte to a specified amount of matrix sample for which an independent estimate of target analyte concentration is available. Spiked samples are used, for example, to determine the effect of the matrix on a method's recovery efficiency.

Red giant (*astronomy*): Old star (spectral type K or M) still performing fusion of hydrogen but having already a helium core.

Reducing atmosphere: Atmosphere with high hydrogen content. Carbon, oxygen and nitrogen mainly exist as CH_4 , H_2O and NH_3 .

Reduction potential: Measure of the tendency for a molecule or an ion to give an electron to an electron acceptor that, itself, can be a molecule, an ion or an electrode. A conventional reduction potential scale for molecules in water allows

to determine, a priori, what chemical species will be the electron acceptor and what will be the electron donor when they are mixed together at a well-defined concentration.

Refractory inclusion: Aggregate of refractory minerals incorporated into a meteorite. See CAI.

Refractory: Substance that remains solid in all temperature conditions available in a particular body of the solar system (e.g., dust particles in a comet). If not refractory, a substance is said to be volatile.

Region (*HII region*): Interstellar cloud such that H exists essentially as H^+ . Ionization is due to intense UV radiation coming from OB stars. HII is the old name used by spectroscopists to describe lines coming from H^+ recombination.

Replication: Biochemical process by which a DNA strand or a RNA molecule is copied into the complementary molecule. Replication is different from transcription and from translation.

Reservoir: In biogeochemistry or geochemistry, a reservoir is the available quantity of one chemical element (or species) within one specific compartment of the global ecosystem (ocean, atmosphere, biosphere, sediments, etc.). Any reservoir may evolve through exchanges with the other ones (biogeochemical or geochemical fluxes).

Residence time: The average time a chemical element or species spends in a geochemical or a biogeochemical reservoir (ocean, atmosphere, biosphere, sediments, etc.). Assuming the reservoir is in steady state (input flux = output = dQ/dt) and contains an amount Q of one element, the residence time $\tau = Q/(dQ/dt)$.

Retrotransposons: DNA sequences able to move from one site to another along the chromosome. Retrotransposons belongs to a transposons family, which requires RNA molecule as an intermediate. The more frequent retrotransposons are the Alu sequences; around one million of such sequences have been identified in the human genome.

Ribose: Aldopentose of major importance in the living world; part of RNA nucleotides.

Ribosomes: Intracellular structures containing many rRNA (ribosomal RNA) molecules together with a complex of 60 to 80 proteins. Synthesis of proteins takes place in ribosomes by condensation of amino acids; information about the correct sequence is given by a mRNA (messenger RNA) while each amino acid is linked to a specific tRNA (transfer RNA).

Ribozyme: RNA molecule acting as an enzyme.

Ridge (*geology*): Submarine mountain chain located at divergent lithospheric plate margins. On the Earth the total length of ridges is 60 000km.

Rift: Rift valley limited by faults = graben.

Rigid rotator (hypothesis): Molecular system frozen at its equilibrium geometry. Interatomic distances and bond angles do not vary during rotation.

Ringwoodite: Mineral, Nesosilicate (isolated SiO_4 tetrahedrons), $[(\text{Mg,Fe})_2\text{SiO}_4]$. Also called olivine γ phase. In the Earth's mantle, ringwoodite is stable between 520 and 660km.

RNA world: Often considered as an early hypothetical stage of life evolution, based on a life without protein and such that RNAs would have played a double role: catalysis and support of genetic information. This theory is mainly based on the discovery of catalytic RNAs (ribozymes) and on an increasing knowledge about the importance of RNAs in contemporary life. For other scientists, the RNA world is the stage of evolution that preceded the DNA emergence.

RNA: Ribonucleic acid, a class of nucleic acids containing ribose as a building block of its nucleotides. RNAs themselves are divided into subgroups like messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA).

Rock: Solid material made up of mineral assemblage. It constitutes telluric planets, asteroids and probably the core of giant planets.

Rocky planets: Other name for telluric planets.

Rodinia: Supercontinent that formed about 750Ma ago.

Root (*genetics*): For a particular genetic tree, the last common ancestor of all the organisms of the tree.

Rosetta: ESA mission launched at the beginning of 2004; it will reach the P/Cheryumov–Gerasimenko comet after a 10-year trip. An orbiter will follow the comet during one year, and a lander will perform in situ analyses on the comet nucleus surface.

Rotatory power: Deviation of the polarization plane of a radiation of well-defined frequency by a chiral medium.

Rovibrational (level): Energy level of a molecule expressed as a rotational quantum number J associated to a given vibration level ν .

r-process (rapid): Nucleosynthetic mechanism that takes place during explosion events (i.e., supernova) and leads to the synthesis of neutron-rich nuclei with atomic number > 26 (iron). During this process, the atomic nucleus captures a huge amount of neutrons and is immediately dissociated through β^- decay.

RRKM: Method of molecular dynamics for the calculation of rate constants of chemical reactions due to Rice, Ramsperger, Kassel and Marcus.

rRNA: Ribosomal RNA. rRNAs are the major components of ribosomes. Some of them have catalytic properties.

Rubisco (or *RuBisCo*): Ribulose-1,5 biphosphate carboxylase-oxygenase; enzyme, which catalyzes the CO_2 metabolism. It is the most abundant enzyme on Earth.

Runaway greenhouse effect: Amplification of a greenhouse effect due to vaporization of molecules able to absorb infrared radiation emitted or received by the planet and which, themselves, contribute to greenhouse effect (i.e., Venus).

Runaway growth (*astronomy*): Increasing rate of planetary accretion with planet growth, it leads to an increasing size difference between the small and large bodies.

R_{sun} : The radius of the Sun (0.69×10^6 km), 1/200 AU, ~ 10 times the radius of Jupiter, ~ 100 times the radius of the Earth.

Runoff channels: Kind of channels observed on Mars and which seem to originate by large water flows over a long period of time.

S

Sagduction: Gravity driven rock deformation. When high density rocks (e.g., komatiites) are emplaced over low density rocks (e.g., TTG), they create an inverse density gradient that results in the vertical sinking of dense rocks in the lighter ones. Sagduction was widespread before 2.5 Ga.

Sarcosine: $(\text{CH}_3)\text{NH-CH}_2\text{-COOH}$, non-proteinic N-methyl amino acid.

Schist: Fine-grained sedimentary rock characterized by a cleavage (slate). It results of fine sediment (i.e., mud) metamorphism.

Secondary structure (*of a protein*): Spatial arrangement of the main chain of a protein. Alpha-helices, beta-sheets, beta-turns are examples of secondary structures.

Sedimentary rock: Rock generated on the Earth's surface. It can consist in the accumulation of rock particles (detrital) or of organic matter (oil, coal). It can also be produced by physicochemical or biogenic precipitation of dissolved ions. Detrital particles and ions derived of alteration and erosion of pre-existing rocks.

Selenocysteine: Frequently described as the 21st proteinic amino acid because it is sometimes coded by the genetic code. It has the structure of the cysteine with a Se atom replacing the S atom.

SELEX (*Systematic Evolution of Ligands by Exponential Enrichment*): Method that generates high-affinity RNA or DNA ligands through successive in vitro rounds of directed molecular evolution. This process allows to identify aptamers by iterative enrichment for molecules capable of binding a target.

Semi-major axis: Orbital parameter. Half of the major axis of an elliptical orbit.

Semi-minor axis: Orbital parameter. Half of the minor axis of an elliptical orbit or small axis of the orbital parameter.

Sense: DNA strand that is not copied into mRNA. Therefore, the sequence of the sense strand corresponds to the mRNA sequence which, itself, corresponds

to the transcription of the antisense strand. Sense and antisense DNA sequences are strictly complementary.

Sequence (*genetics*): Series of directly linked nucleotides in a DNA or a RNA strand; series of directly linked amino acids residues in proteins.

Sequence (main sequence) (*astronomy*): Stage in star evolution when hydrogen fusion takes place in the core. In the HR diagram, main sequence stars draw a straight line.

Sequencing: Experimental determination of a DNA, RNA or protein sequence.

Serine (*Ser*): Proteinic amino acid with a $-\text{CH}_2\text{-OH}$ side chain. Being hydrophilic, serine is generally present in the external part of the skin and can be used as the proof for human contamination of meteoritic samples. Serine is produced in very small amounts during simulation experiments considered as experimental models of interstellar chemistry.

Serpentine: Mineral family, Phyllosilicate (water-bearing sheet silicate) (i.e., Antigorite $[\text{Si}_4\text{O}_{10}\text{Mg}_6(\text{OH})_8]$). They are generated by olivine (and sometimes pyroxene) alteration. They play an important role in the internal water cycle.

SETH (*Search for Extraterrestrial Homochirality*): Search for proofs of enantiomeric excesses in extraterrestrial objects.

SETI (*Search for Extraterrestrial Intelligence*): Search for electromagnetic signals (essentially in the radiowave domain) that should be intentionally emitted by some living organisms and coming from sources located outside the solar system.

Shield (*geology*): Huge block of old (often Precambrian in age) and very stable continental crust, e.g., Baltic shield (includes Finland, Sweden, Norway, and Western part of Russia).

Shocked quartz: Quartz showing characteristic microscopic defects of its crystalline structure that were generated by the passage of a high pressure (> 5 Gpa) shock wave. The most common modification is the formation of planar deformation features (PDF) consisting of fine lamellae of amorphous SiO_2 oriented parallel to two (or more) specific crystallographic planes. Under even higher pressure the whole quartz crystal can be transformed to glass. Shocked quartz are a diagnostic criteria to recognize meteorite impacts as no other natural process is capable of producing the required high-pressure dynamic shock wave.

Siderite: Mineral, iron carbonate $[\text{FeCO}_3]$.

Siderophile: Elements frequently associated to iron (like Au, Pt, Pd, Ni, etc.).

SIDP (*Stratospheric IDP*): IDP collected in the stratosphere by captors placed on airplanes.

Silicate: Wide mineral family of silicium oxides. The structure is based on a tetrahedron $(\text{SiO}_4)^{4-}$.

Simple sugar: Old name for monosaccharide.

Site (*active site*): For an enzyme, specific locus where the substrate is fixed, ready to react with the reactant.

SL (*astronomy*): Solar luminosity (3.826×10^{24} W) or 3.826×10^{33} erg s⁻¹.

SM: Solar mass (2×10^{30} kg).

Small bodies: Comets and asteroids.

SMOW (*Standard Mean Ocean Water*): Standard reference sample for H and O isotopic abundance measurements.

SNC: Family of about 30 meteorites, which, based on several experimental observations, are considered to have a Martian origin. SNC is the abbreviation for Shergotty, Nakhla and Chassigny, the first three meteorites of this family.

Snow line: Limit between the regions where H₂O is gaseous and the region where it is solid. In a protoplanetary disk, the snow line marks the boundary between the region where telluric planets form from silicates, and the region where giant planets, made of an icy core surrounded by protoplanetary gases, are generated.

Snowball effect (*astronomy*): During planetary accretion the larger objects grow more rapidly than the smaller ones, leading to an increasing difference in size between small and large bodies. This process accelerates (like a rolling snowball, the growing rate is proportional to its size), until the moment when the larger objects control the whole dynamic. See also runaway growth.

Solar constant: Total energy delivered each second by the Sun and measured in W m⁻², on a surface being placed at 1 AU of the Sun and perpendicularly to the light rays. The value of the solar constant is equal to 1360 W m⁻².

Solar type star: Star of G type, similar to the Sun.

Solidus: Line, which, in composition versus temperature or pressure versus temperature diagrams, separates the domain where crystals and liquid coexist from the field where only crystals (solid) exists.

Spallation: Atomic nuclei breaking due to the collision between two atomic nuclei with energy greater than the Coulombic barrier; it leads to the formation of new elements, stable or radioactive.

Speckle interferometry: Acquisition and statistical treatment of images technique that recovers part of the information included in astrophysical images and lost because of the atmosphere's imperfect transmission. This technique, developed during the 1980s is rarely used and is replaced by adaptive optics techniques that allow a real time partial correction of atmospheric turbulence effects.

Spectral type: Star classification procedure based on electromagnetic spectrum that, itself, depends on the star's surface temperature. The OBAFGKM classification ranges from surface temperature of about 50 000 K (O type) to 3500 K (M type).

Spinel: Mineral, Oxide $[\text{MgAl}_2\text{O}_4]$, Stable at low pressure (depth $< 70\text{km}$). In the mantle, at shallow depths, spinel is the only aluminum-bearing phase.

Spore (*biology*): (1) Resistant, dormant, encapsulated body formed by certain bacteria in response to adverse environmental conditions. (2) Small, usually single-cell body, highly resistant to desiccation and heat and able to develop into a new organism.

s-process (slow): Nucleosynthetic mechanism that takes place over long periods of time in environments with moderated neutron flux (i.e., red giant). This mechanism leads to the synthesis of atoms with atomic number > 26 (iron) that are located in the nuclear stability valley.

Star: Celestial object, generally spherical, where thermonuclear reactions take place (e.g., the Sun) or will take place in the future (e.g., the PMS stars) or has taken place in the past (neutron star).

Stellar cluster: Group of a few hundred to several million stars. The smallest groups are named associations. Most of the stars are formed in open clusters.

Stereoisomers (*or stereomers*): Isomers (molecules with identical atomic composition, i.e., the same number of the same atoms) such that the atoms are identically interconnected by covalent bonds. If two stereoisomers are different in the same way that a left hand is different from a right hand, the stereoisomers are enantiomers. In all other cases, stereoisomers are called diastereoisomers (or diastereomers).

Stereoselectivity: When a chemical reaction leads to the formation of stereoisomers and when the latter are not exactly produced in the same amount, the reaction is called stereoselective. Similarly, when two stereoisomers react at a different rate with a reactant, the reaction is stereoselective. In the chemical literature, some authors have introduced subtle differences between stereoselectivity and stereospecificity.

Steric effect: One of the multiple effects introduced by organic chemists to explain the relative stability or the relative reactivity inside a group of molecules. The steric effect takes into account the size of each atom or groups of atoms. For instance, the steric effect of a $-\text{C}(\text{CH}_3)_3$ group is larger than the steric effect of a CH_3 group. The steric effect can be explained on the basis of repulsive term in the van der Waals forces.

Stony meteorites: Mainly made up of silicates, they can contain from 0 to 30% of metal grains and several percents of sulfides. They can be differentiated (achondrite) or undifferentiated (chondrite) even if CI undifferentiated chondrites do not contain chondrules.

Stop codon (*genetics*): Codon that does not code for amino acids but that indicates the translation end. For the Universal code, these codons are UAA, UAG and UGA (synonym: termination codons).

Stratopause: Atmospheric boundary between stratosphere and mesosphere.

Stratosphere: Atmospheric layer located above the troposphere and below the mesosphere, between 9–17km and 50km. In the stratosphere, temperature slightly increases with altitude preventing convective movements.

Strecker (*synthesis*): Synthetic method producing amino acids from aldehyde, HCN, NH₃ and water. Frequently considered as important prebiotic reaction.

Stromatolite: Sedimentary structure consisting of laminated carbonate or silicate rocks and formed by the trapping, binding, or precipitating of sediment by colonies of microorganisms (bacteria and cyanobacteria).

Strong force: One of the four fundamental forces in physics that contribute to the stability of the atomic nucleus.

Subduction: Plate tectonic mechanism, where an oceanic plate sinks under an other lithospheric plate (generally a continental plate, but sometimes also an oceanic plate).

Sublimation: Direct phase change from solid to gas state.

Succinic acid: HOOC-CH₂-CH₂-COOH.

Suevite: Rock associated with impact craters (impactite). Defined in the Ries crater (Bavaria), this brecciated rock containing melt material (glass) consists of fragments of impacted lithologies and basement rocks, mixed in a fine-grained clastic matrix.

Sun: Star belonging to the main sequence (in HR diagram); it is 4.56 Ga old. Sun–Earth distance corresponds to 8 light-minutes or 1 astronomical unit (1 AU).

Supernova: Exploding star, which, before the explosion, was either a binary star (type I) or a massive star (type II). After explosion, the remnant becomes a neutron star.

Surface Plasmon Resonance (SPR): A biosensor system used for analyzing ligand binding and kinetics of specific molecules within complex mixtures without prior purification. Binding of a ligand to a biomolecule immobilized on a membrane (metallic film) results in changes in membrane surface plasmon resonance.

Symbiosis: Prolonged association between two (or more) organisms that may benefit each other. In the case of endosymbiosis, one organism lives into another one; their relationship is irreversible and implies a complete interdependence such as the two become a single functional organism. Mitochondria and chloroplasts are remnants of the endosymbiosis of photosynthetic bacteria.

Synchrotron radiation: Electromagnetic waves covering a large frequency domain and emitted under vacuum by high velocity electrons or ions, when their trajectory is altered, as by a magnetic field. Synchrotron radiation is naturally polarized.

Synonymous (*codons*): Different codons that specify the same amino acid, e.g., AAA and AAG specify lysine.

T

T Tauri star: New-born low mass star (lower than two solar masses) starting to become optically observable. Classical T Tauri are very young (less than 1 Ma old) and are not yet on the main sequence (PMS stars). They are surrounded by a disk and their luminosity is variable with an excess of IR with respect to UV. T Tauri stars we observe today are probably similar to the young Sun.

Talc: Mineral, Phyllosilicate (water-bearing sheet silicate) $[\text{Mg}_3(\text{Si}_4\text{O}_{10})(\text{OH})_2]$. Talc frequently results of hydrous alteration of magnesian minerals (i.e., olivine).

Taxonomy: Science that classifies living species. Similar species belong to the same taxon.

Tectonic: Science of rock and crustal deformation.

Tektites: Si-rich, natural glasses formed by melting of the upper part of the target lithologies during a meteorite impact. Tektites are ejected from the crater and spread over a broad geographic area called a Strewnfield.

Telluric planets: Small and dense rocky planets (density: 3 to 5.5 g cm^{-3}). These planets (namely Mercury, Venus, Earth and Mars) are silicate-rich and were formed in the inner part of the protosolar nebula, beyond the dust line.

Terraforming: Voluntary transformation of a planetary atmosphere in order to allow colonization by plants and animals (including man).

Tertiary structure (of a protein): Arrangement of the side chains of the residues in space, overall conformation of a protein.

Theia: Name given by some authors to Mars-sized object, which, after impacting the young Earth, led to the Moon formation. This cataclysmic event took place 4.5 to 4.45 billion years ago.

Thermolysis: Thermal degradation of a molecule into smaller fragments (atoms, radicals, molecules and, rarely, ions).

Thermonatrite: Mineral, Water-bearing sodium carbonate $[\text{Na}_2\text{CO}_3\text{H}_2\text{O}]$. This mineral is water-soluble and primarily forms in evaporite deposits and in desert soils where it may occur as a surface deposit, or in volcanic fumaroles.

Thermonuclear reaction: Reaction leading to formation of heavier nuclei by fusion of lighter nuclei. Thermonuclear reactions require high T and high P conditions. In natural conditions, these reactions occur spontaneously in cores of main sequence stars or of heavier stars like giant stars. It occurs also during supernovae explosions.

Thermopause: Atmospheric boundary between thermosphere and exosphere.

Thermophile: Organism living optimally at high temperatures. They can be divided in moderate thermophiles, living optimally between 40 and 60°C , extreme thermophiles, between 60 and 80°C , and hyperthermophiles, living optimally at temperatures higher than 80°C (and up to $\sim 120^\circ\text{C}$).

Thermosphere: Atmospheric layer located above the mesosphere and below the exosphere. At the top of the thermosphere, at an altitude of about 700 km (on Earth) the temperature can be $> 1300\text{K}$.

Thin section (*geology*): Rock slice generally $30\mu\text{m}$ thick. At such a thickness most minerals are transparent and can be observed by transmitted light with a polarizing microscope.

Thiocyanate: R-S-CN.

Thioester: R-S-CO-R'.

Thiol: R-SH.

Tholeiite: Relatively silica-rich and alkali-poor basalt.

Tholins: Solid mixture of complex organic molecules obtained by irradiation of reduced gases like CH_4 , NH_3 . Could be present on Titan.

Threonine (*Thr*): Proteinic amino acid containing four C atoms and a -OH group in its side chain. Threonine is described as a hydrophilic amino acid.

Thymine (*T*): Nucleic base belonging to purine and specific of DNA.

Titan: The biggest satellite of Saturn ($\sim 5000\text{km}$ in diameter; approximately the same size as Mercury). In the solar system, this is the only satellite possessing a dense atmosphere. Very probably, complex organic reactions could have taken place in its multicomponent hydrocarbon-bearing atmosphere.

Titus–Bode law: Empirical law giving approximately the planet–Sun distance d as a function of the planet ranking n ($d = (4 + 3.2^{(n-2)})/10$). Mercury is an exception to this law ($d = 0.4\text{AU}$).

Tonalite: Plutonic magmatic rock (granitoid), made up of quartz and calcic plagioclase feldspar; biotite and sometimes amphibole are minor mineral phases. Tonalite does not contain alkali feldspar. Dacite is its effusive equivalent.

TPF (*Terrestrial Planet Finder*): NASA project with a similar goal as the ESA Darwin project, i.e., discovery of extrasolar terrestrial planets.

Transcription: Synthesis of an mRNA as a copy of an antisense DNA single strand.

Transduction: Transfer of genetic material from one bacteria to another through viral infection.

Transfer of genes (*horizontal transfer*): Transfer of a gene from one organism to another that does not belong to the same species. Such transfer can occur through viral infection or by direct inclusion of genetic material present in the external medium. Horizontal transfer is different from vertical transfer from parents to children.

Transferase: Enzyme that catalyzes the transfer of a chemical group from one substrate to another.

Transform fault: Boundary between two lithospheric plates, which slide without any crust creation or destruction.

Transit: Motion of a planet in front of the disk of its star.

Translation: Sequence enzymatic reactions such that the genetic information coded into a messenger RNA (mRNA) leads to the synthesis of a specific protein.

Transneptunian object (TNO): Solar system body located beyond the orbit of Neptune, in the Kuiper belt or beyond.

Triangular diagram: Diagram classically used in geology to plot the chemical or mineralogical composition of a sample. Each apex of the triangle represents 100% of one component whereas the opposite side corresponds to 0% of the same component.

Triple point: In a P - T phase diagram of a pure compound, it corresponds to the unique P , T value where the three phases (gas, liquid, solid) coexist at equilibrium. For water, $P = 6.11$ mbar and $T = 273.16$ K.

tRNA synthetase: Enzyme catalyzing, in a very specific way, bond formation between an amino acid and its tRNA.

tRNA: Transfer RNA.

tRNA: Transfer RNA. Polymer containing 70 to 80 ribonucleotides and specific of each amino acid to which it is linked. Able to recognize a triplet of nucleotides of mRNA (codon) by a specific molecular recognition process involving a triplet of nucleotides of the tRNA (anticodon). The tRNAs play a fundamental role for proteinic synthesis.

Trojans: Family of asteroids located at the Lagrange point on the Jupiter orbit. Their position together with that of the Sun and Jupiter determines an equilateral triangle.

Trondhjemite: Plutonic magmatic rock (granitoid), made up of quartz and sodic plagioclase feldspar; biotite is a minor mineral phase. Tonalite and trondhjemite are similar rocks except that tonalite contain a calcic plagioclase whereas it is sodic in trondhjemite).

Tropopause: Atmospheric boundary between troposphere and stratosphere.

Troposphere: Lowest part of the Earth's atmosphere, as the temperature at its base is greater than at its top, it is a place of active convection. On Earth, troposphere thickness ranges between 9km (pole) and 17km (equator). Most meteorological phenomena take place in troposphere.

Tryptophane (Trp): Proteinic amino acid containing eleven C atoms; tryptophane contains a heterocycle in its side chain. It is described as an aromatic amino acid.

TTG: Tonalite, Trondhjemite, Granodiorite. Rock association typical of the continental crust generated during the first half of the Earth's history.

Tunguska event: Explosion that took place in (June 30, 1908) Siberia and devastated 2000km² of forest. It was probably due to the explosion in the atmosphere of small (< 50m) asteroid or comet.

Tunneling effect: Description of tunneling effect requires the use of quantum mechanics because it is a direct consequence of the wave properties of particles. When a system A gives another system B while the internal energy of A is lower than the energy barrier (activation energy) required to cross the barrier from A to B, it can be said that A gives B by passing through the barrier by a tunnel effect.

Turbidite: Sedimentary deposit deposited by a turbidity current according to a characteristic fining upward grain size sequence in deep water at the base of the slope and in the abyssal plain.

Turbidity current: Also called density current. Viscous mass of mixed water and sediment that propagates downward along the continental slope (of an ocean or a lake) due to its greater density. It may reach high speeds and has a high erosive power. Such currents are set in motion by earthquakes, for example.

Turbulence (parameter) (*astronomy*): In hydrodynamics, the α model empirically describes the turbulent viscosity of a flow. It is based on a parameter α : $\nu_t = \alpha \frac{C_s^2}{\Omega_s}$ with C_s = local sound speed and Ω_s the Keplerian rotational frequency.

Tyrosine (*Tyr*): Proteinic amino acid containing nine C atoms. Its side chain contains a phenolic group.

U

Upwelling: Upward movement of cool and nutrient-rich subsurface seawaters towards the surface. Upwelling is mainly controlled by local atmosphere dynamics (pressure, wind), frontal contact between water masses with different densities (oceanographic fronts) or by Eckman transport.

Uracil (*U*): Nucleic base belonging to pyrimidine family and specific of RNAs.

Urea: (H₂N-CO-NH₂) First organic molecule that has been synthesized from a mixture of inorganic molecules (Wohler).

UV radiation: Electromagnetic radiation characterized by wavelengths ranging from 0.01 to 0.4 microns (energies from 124 to 3.1 eV).

V

Valine (*Val*): Hydrophobic proteinic amino acid containing five carbon atoms.

Van der Waals forces (*VdW*): Interatomic forces acting between non-bonded atoms at the intramolecular level but also at the intermolecular levels. Repulsive VdW forces are responsible for the non-infinite compressibility of matter and for the fact that atoms have sizes. Attractive VdW forces are responsible for matter cohesion. VdW forces play a major role in biochemistry: together with H-bonds

and electrostatic interactions, they determine the preferred conformations of molecules and they contribute to molecular recognition phenomena.

Vernal point: Sun location on the celestial sphere at the vernal equinox (spring equinox). It is the origin of coordinates in the equatorial system.

Vertical tectonic: See sagduction.

Vesicle (*geology*): Bubble-shaped cavities in volcanic rocks formed by expansion of gas dissolved in the magma.

Vesicle: Small sac, made of hydrophobic or amphiphile molecules, whose content is isolated from the surrounding environment.

Viking: NASA mission to Mars, which started in 1976. The two landers (Chryse Planitia and Utopia Planitia) performed a series of very ambitious experiments to detect the presence of life on Mars. Unfortunately, many results were ambiguous.

Virus: System containing DNA or RNA surrounded by a proteinic envelope (capside). When introduced in a living cell, a virus is able to replicate its genetic material by using the host cell machinery.

VLBI (*Very Long Baseline Interferometry*): Technique that allows a very accurate determination (50 microarcsec) of the position of astronomical sources of radiowaves. This method is based on interferometry measurements using very distant radiotelescopes (large base) located on the same continent or on different continents or even on Earth and on a satellite.

VLT (*Very Large Telescope*): Group of four large telescopes (4 to 8 meters) and several smaller telescopes, able to work as interferometers and located in Chile. VLT is managed by ESO (European Southern Observatory).

Volatile (*volatile substance*): Molecule or atom that sublimates at relatively low temperatures (i.e., cometary ices).

Volatile: See refractory.

Volcanoclastic sediment: Sediment due to sedimentation in the sea or in a lake of volcanic products (i.e., ashes).

W

Wadsleyite: Mineral, Nesosilicate (isolated SiO_4 tetrahedrons), $[(\text{Mg,Fe})_2\text{SiO}_4]$. Also called olivine β phase. In the Earth's mantle, wadsleyite is stable between 410 and 520km.

Wall (*of a cell*): Extracellular membrane. In bacteria, cell wall structure is complex: the walls of Gram positive and Gram F negative bacteria are different.

Water triple point: See triple point.

Watson–Crick: Canonical model of DNA (double helix) involving the pairing of two polynucleotide strands via H-bonds between A and T or G and C. RNA is generally single-stranded but within a single strand Watson–Crick pairing can occur locally. When it happens, it involves AU and GC pairing.

Weak bonds: Intermolecular or intramolecular bonds involving non-bonded atoms (atoms not bonded by covalent, coordination or electrostatic bonds). H-bonds are well-known examples of weak bonds but van der Waals forces and electrostatic interactions also contribute to weak bonds. The weak bonds play a fundamental role in the living world: they determine the conformation of molecules and more particularly the conformation(s) of biopolymers; they are responsible for the specificity of molecular recognition. The intermolecular association due to weak bonds is generally reversible.

Weak force: One of the four fundamental forces of physics. Parity is violated for these forces. The coupling between weak forces and electromagnetic forces is at the origin of the very small energy difference between enantiomers (PVED for Parity Violation Energy Difference).

Weathering: See alteration.

White dwarf: Relic of a dead star with an initial mass $< \sim 8 M_{\text{Sun}}$. Its gravity collapse is limited by electron degeneracy pressure.

Wind (*solar or stellar*): Flow of ionized matter ejected at high velocity (around 400km/s) by a star. Solar wind mainly contains protons.

Wobble (*genetics*): Describes imprecision in base pairing between codons and anticodons. It always involves position 3 of codon, mainly when the base is U.

X

Xenolith: Inclusion or enclave of foreign rock or mineral (xenocrystal), in a magmatic rock.

Y

Young sun paradox: Apparent contradiction between the lower brightness of the young Sun, (70% of the present-day intensity in the visible spectral range) and the early presence of liquid water (-4.4Ga) on Earth. A strong greenhouse effect due to high concentrations of atmospheric CO_2 could account for this apparent paradox.

YSO (*Young Stellar Object*): Star which has not yet completed the process of star formation. YSO includes objects ranging from dense cores, (that can be detected in the submillimeter IR frequency range) to pre-main sequence stars (T Tauri, Herbig AeBe) and HII regions.

Z

Z (*astronomy*): Abundance of heavy elements, i.e., all elements except H and He.

ZAMS (*Zero Age Main Sequence*): Ensemble of new-born stars in which H fusion has just started.

Zircon: Mineral, Nesosilicate (isolated SiO_4 tetrahedrons), $[\text{ZrSiO}_4]$. This mineral, which also contains traces of Th and U, is extremely resistant to weathering and alteration. These are the reasons why it is commonly used to date rock. The oldest zircon crystals so far dated yielded an age of 4.4Ga. They represent the oldest known terrestrial material.

Zodiacal light: Diffuse faint light observed in a clear sky close to the ecliptic. It is due to the diffusion of the solar radiation by the electrons and the interplanetary dust. Also used for any light diffused by dust particles in a planetary system.

Authors (Photos and addresses)

ALBAREDE Francis

Professor
Ecole Normale Supérieure
Lyon, France
E-mail: albarede@ens-lyon.fr

Geochemistry: early Earth



BENILAN Yves

Research Scientist CNRS
Laboratoire Interuniversitaire
des Systèmes Atmosphériques
Université Paris 12, France
E-mail: benilan@lisa.univ-paris12.fr

Planetology: space science and exobiology

BERSINI Hugues

Professor
Director of IRIDIA
(Institut de Recherche Interdisciplinaire
et de Développements en Intelligence Artificielle)
Université Libre de Bruxelles, Brussels, Belgium
E-mail: bersini@ulb.ac.be

Informatics: artificial intelligence and life, oriented
object informatics, bioinformatics





BERTRAND Philippe

Research Director CNRS
Université Bordeaux 1, Bordeaux, France
E-mail: p.bertrand@epoc.u-bordeaux1.fr

**Paleoclimatology/Paleoceanography/
Biogeochemistry**

CHAUSSIDON Marc

Research Director CNRS
Centre de Recherches Pétrographiques
et Géochimiques (CRPG), Nancy
E-mail: chocho@crpg.cnrs-nancy.fr

Geochemistry/Cosmochemistry: early solar system evolution, isotopic composition of meteorites and archeans rocks



CLAEYS Philippe

Professor
Vrije Universiteit Brussel, Department of Geology
Pleinlaan 2, 1050 Brussels, Belgium
E-mail: phclaeys@vub.ac.be

Geology: comets and asteroids impacts and their consequences for life evolution

COTTIN Hervé

Assistant Professor
Laboratoire Interuniversitaire
des Systèmes Atmosphériques
Université Paris 12, France
E-mail: cottin@lisa.univ-paris12.fr

Astrochemistry: spatial organic physicochemistry, comets and interstellar medium chemistry



ELLINGER Yves

Research Director CNRS
 Laboratoire d'Etude Théorique
 des Milieux Extrêmes (LETMEX)
 et Laboratoire de Chimie Théorique (LCT)
 Université Pierre et Marie Curie, Paris, France
 E-mail: ellinger@lct.jussieu.fr

Chemistry: unstable species structure, solid-gas interface, potential surface calculations

**ENCRENAZ Thérèse**

Research Director CNRS
 Laboratoire d'Etudes Spatiales
 et d'Instrumentation pour l'Astrophysique
 Observatoire de Paris-Meudon, France
 E-mail: therese.encrenaz@obspm.fr

Planetology: solar system: atmospheres of planets, satellites and comets (chemical composition, thermal structure). Origin and evolution of solar-system bodies. Remote sensing techniques: infrared and millimeter spectroscopy of planetary atmospheres

FORGET François

Research Scientist CNRS
 LMD, Institut Pierre-Simon Laplace, Paris, France
 E-mail: forget@lmd.jussieu.fr

Planetary Atmosphere: Mars present and past climate modeling

**GARGAUD Muriel**

Research Scientist CNRS
 Observatoire Aquitain des Sciences de l'Univers
 Université Bordeaux 1, Bordeaux, France
 E-mail: muriel@obs.u-bordeaux1.fr

Astrophysics/Astrobiology: interstellar medium physicochemistry, origins of life



LOPEZ-GARCIA Purificación

Research Scientist CNRS
Unité d'Ecologie, Systématique et Evolution
Université Paris-Sud, France
E-mail: puri.lopez@ese.u-psud.fr

Microbiology: microbial diversity, microbial evolution, extreme environments

MARTIN Hervé

Professor
Laboratoire Magmas et Volcans
Université Blaise Pascal, Clermont-Ferrand, France
E-mail: h.martin@opgc.univ-bpclermont.fr

Geochemistry: geological and geochemical evolution of the primitive Earth. Subduction zone magmatism



MOREIRA David

Research Scientist CNRS
Unité d'Ecologie, Systématique et Evolution
Université Paris-Sud, France
E-mail: david.moreira@ese.u-psud.fr

Biology: molecular phylogeny, microbial evolution

OLLIVIER Marc

Assistant Astronomer
Institut d'Astrophysique Spatiale
Université Paris-Sud, Orsay, France
E-mail: marc.ollivier@ias.u-psud.fr

Astronomy: extrasolar planets, remote detection of life



PAUZAT Françoise

Research Director CNRS
Laboratoire d'Etude Théorique
des Milieux Extrêmes (LETMEX)
et Laboratoire de Chimie Théorique (LCT)
Université Pierre et Marie Curie, Paris, France
E-mail: pauzat@lct.jussieu.fr

Astrochemistry: interstellar medium and planetary atmospheres physicochemistry, numerical simulation of spectral signatures and of elementary chemical processes

**PRANTZOS Nikos**

Research Director CNRS
Institut d'Astrophysique de Paris, France
E-mail: prantzos@iap.fr

Astrophysics: nuclear astrophysics, high-energy astronomy, galaxy evolution

PRIEUR Daniel

Professor
Laboratoire de Microbiologie
des Environnements Extrêmes
Université de Bretagne Occidentale, Brest, France
E-mail: daniel.prieur@univ-brest.fr

Microbiology

**SELSIS Franck**

Research Scientist CNRS
Centre de Recherche Astronomique de Lyon
Ecole Normale Supérieure de Lyon, France
E-mail: franck.selsis@ens-lyon.fr

Planetology: formation and evolution of planetary atmospheres, extrasolar planets, biomarkers

Index

- Acasta 83
- accretion 131
- acetogenic 311
- actin fiber 226
- age
 - geochemical 47
- albedo 201
- Alvinella pompejana* 327
- amino acid 351, 358
- anaerobe 230
- anorthosite 92
- anoxygenic 231
- Archaea 226, 305, 329
- Archean 267, 270, 272
- Arnon cycle 311
- artificial life 223, 491, 508
- asteroid 239, 243, 264, 272
- astrometry 182
- atmosphere 123
 - CO₂ 291
 - gas signature 188
 - thermodynamic escape 179
- ATP 230, 311
- ATPase 311
- attractor 514–517
- autocatalysis 222
- autopoiesis 222, 503, 533
- autoreplication 222

- Bacteria 223, 305, 329
- Bathymodiolus* 326
- bicarbonate 298
- bio
 - chemistry 229, 305
 - geochemistry 282
 - sphere 221
 - synthesis 312
- bipolar 230

- black body emission 551
- black smokers 354
- Bracewell interferometer 191, 560
- breccia 244, 255
- brown dwarf 147, 549

- CAIs 51–52
- Callisto 356
- Calvin–Benson cycle 311
- Calyptogenia magnifica* 325
- capsid 223
- carbon 229
 - dioxide cloud 207
 - fixation 310
 - inorganic 289
 - ocean buffering 296
 - ocean reservoir 300
 - reservoir 297, 298
- carbonate
 - bioprecipitation 294
 - carbonate-silicate cycle 206, 215
 - compensation depth 295
 - dissolution 294
- carbonates 411
- carbonic ice 105, 118
- catalysis 308
- cell
 - functions 224
 - nucleus 226
 - nutrition 224
 - plasma membrane 225, 306
 - replication 225
 - reproduction 225
 - respiration 230, 294
- cellular automata 496, 529
- cenancestor 305
- chemical evolution 352
- chemical reaction network 500, 512, 521

- chemo
 - lithoautotroph 310
 - synthesis 230
- chemoton 505
- Chicxulub 242, 246, 254, 255, 258–260, 262, 264, 270
- chloroplast 226
- chondrite 51, 351
- chromosome 225
- chronologic calibration 58
- collision 239, 241–243, 267, 268, 272
- combustion
 - carbon 17, 18, 272
 - helium 12, 13
 - hydrogen 10, 11
 - neon 19
 - oxygen 20
- comet 123, 239, 243, 264, 272, 349, 356
 - experimental simulation 358
 - Shoemaker Levy 239
- condensation 51, 52
- coronagraph 190
- COROT 184
- COSAC 378
- covalent link 229
- Cretaceous-Tertiary (KT) boundary 242, 254–256, 258–264, 266, 267, 270
- crust
 - continental 76, 80–83
 - oceanic 76, 81
- cumulate 93
- cyanobacteria 230, 311
- cytoplasm 225
- cytoskeleton 226

- Darwin 201, 305
- DARWIN/TPF 191, 193, 216
- denitrification 288
- diazotrophy 288
- DNA 223, 306
- DNA polymerase 309
- Doppler–Fizeau effect 555
- dormant state 222
- dust
 - devil 106
 - storm 105, 106, 108
- ecosystem 233
- ejecta 254, 257, 259, 264, 266, 268

- electron donor 232, 311
- ELODIE 161, 181
- endo
 - plasmic reticulum 226
 - symbiosis theory 285
 - symbiotic 226
- energy metabolism 223
- enzyme 312
 - recruitment 310
- Epicurus 157
- escape velocity 128
- ether linkage 311
- Eucarya 226, 305
- eukaryote 225, 308
- Europa 215, 216, 355
- exergonic 311
- exobiological experiment 407
- exonuclease 309
- exoplanet 147, 212, 216
 - atmosphere 175
 - detection 186
 - eccentricity 172
 - spectrum 175
- extinct radioactivity 47
- extinction 242, 254, 266, 270, 272
- extrasolar planet 212
- extreme environment 233
- extremophile 233

- fatty acid 228, 311
- fermentation 231, 310
- functional emergence 495

- GAIA 165, 182
- Galileo 355
- game of life 530–532
- gamete 225
- Ganymede 356
- gene 308
 - duplication 310
 - genetic algorithm 496, 537
 - genetic code 305
 - genotype 306
 - horizontal transfer 308
- GENIE 188
- genome 223, 309
- genomics 308
- giant planet 213, 214
 - atmosphere 175

- spectrum 175
- Giotto 367
- glacial-interglacial fluctuations 291
- glacier 112
- global ecosystem 284, 291
- greenhouse 125, 202
- Gullies 113

- habitable zone 148, 201, 221
 - continuously habitable zone 210
 - galactic habitable zone 212
 - habitability 199, 221
- Hadean 268, 272
- Hadley's circulation 126
- Hamiltonian 430
- HARPS 181
- HD 209458 b 178
- hematite 116
- Hertzsprung–Russel (HR) diagram 553
 - main sequence 159, 553
- heterosphere 124
- heterotrophs 231, 310
- Hill radius 165
- Hipparcos 160
- homolytic 432
- homopause 124
- hot Jupiter 212
- Huygens 158
- Huygens probe 396
- hydrogenoid 436
- hydrostatic pressure 335
- hydrothermal
 - activity 320
 - vent 354
- hydroxypropionate pathway 311
- hypertelescope 193
- hyperthermophile 234, 312, 332

- immune network 512
- impact 239
 - crater 241, 242, 244, 249, 250, 266, 267
 - deep impact 371
 - impactite 244, 255
 - structure 240
- interferometer 188
 - Bracewell 191, 560
- interstellar grain 349
- ionizing radiation 337
- Iridium (Ir) anomaly 254, 260

- irradiation 66
- isochron 48
- isoprenoid 228, 311
- Isua 83

- Jack Hills 83
- Jean's evaporation 177
- JIMO 356
- Jupiter 214
 - hot Jupiter 212

- Keck-I 182
- Kepler 185
 - laws 550
- KT (Cretaceous–Tertiary) boundary 242, 254–256, 258–264, 266, 267, 270
- Kuiper Belt 132, 357

- last common ancestor 305
- Late Heavy Bombardment 215, 268, 269, 272

- law
 - Kepler 550
 - Planck 551
 - Stefan–Boltzmann 551, 552
- LBT 181
- LINC 181
- lipid 223, 306
- lithosphere 76
- LUCA 305

- macromolecule 228
- magnetosphere 130
- main sequence 159, 553
- Mars 103, 207, 398, 399
 - climate 103
 - dust cycle 105, 108
 - early Mars 206, 208
 - exploration rover 116
 - Express 107, 116
 - Global Surveyor 105, 108, 114, 116, 117
 - obliquity 111
 - Odyssey 108, 112
 - polar caps 105
- mass extinction 242, 254, 266, 270, 272
- melted rock 252, 255, 259, 273
- mesosphere 124
- metabolism 309, 310

- metabolite 230
- metallicity 210, 212
- meteorite 351
 - Murchison 351
- methane 207, 208
- methanogen 208, 231, 311
- microfossil 314
- micrometeorite 352
- microorganism 235
- microtektite 253
- Miller 352
- mitochondria 226
- molecular biology 305
- molecular cloud 349
- Moon 239, 242, 243, 246, 268, 269
- mutation 222, 309

- N-biofixation 288
- NACO 188
- NAD(P)H 232
- network
 - chemical reaction 500, 512, 521
 - neural 512
 - Valley networks 115
- Newton 550
- nitrate 288
- nitrogen 287, 288
- nucleic acid 223
- nutrient 289
- nutrition 224

- ocean
 - magmatic 84, 86, 95
 - planet 180, 215
- OGLE 182
- oligoelements 229
- Oort Cloud 357
- Opportunity 116
- organic matter 347
- outflow channel 113
- oxygen 230, 231, 282
 - minimum zone 288
 - oxygenic photosynthesize 311

- panspermia 200, 222
- parasite 223
- pegaside 176
- permafrost 233
- Phanerozoic 266, 267, 273, 284

- phenotype 306
- phosphate 287
 - residence time 287, 288
- phospholipid 228, 311
- photo
 - autotroph 231
 - synthesis 230, 294
 - trophy 231
- phylogeny 226, 308
- Planck law 551
- planet 123, 549
 - embryo 214
 - extrasolar 212
 - formation 214
 - giant planet 181, 213, 214
 - nucleation 131
 - ocean planet 180, 215
 - Planet finder 188
 - spectral classification 176
 - star – planet distance distribution 162
 - transit 182, 183
- plasma membrane 225, 306
- polymerase 223
- polysaccharide 228
- pre-cell theory 306
- prebiotic soup 310
- PRIMA 182
- progenote 306
- prokaryote 225, 308
- protein 223, 306
- Proterozoic 246, 266
- proto-planetary disk 131
- proton gradient 306
- proton-pump 312
- protosolar cloud 131
- pyrite 310

- quartz 250, 251, 255
 - shocked 255, 258, 266, 267

- redox reaction 230
- reduction power 232
- regulation 281
- reinforcement learning 539
- ribosome 223, 306
- ribozyme 308
- Riftia pachyptila* 325
- rigid rotor approximation 458

- Rimicaris* 328
RNA 223
 messenger RNA 306
 polymerase 308
 world 308
Roche lobe 168
Rosetta 371
rovibrational 431
RRKM 475
- S layer 225
sedimentary carbon reservoir 299
self-replication 503
serpentine 87–88
shatter cone 250
shocked quartz 255, 258, 266, 267
Shoemaker Levy 239
signal recognition particle, SRP 312
SIM 165, 182
snow line 132
Solar System 550
 accretion 131
 age 45
 nebula 349
 wind 130
space missions 347, 360–376, 408
speciation 312
spectrum of extrasolar planet 175
spherule 258, 259, 266, 267, 270
spore 225
Stardust 371
Stefan–Boltzmann law 551, 552
- stereochemistry 311
stratosphere 124
suevite 244, 255, 256
symbiosis 227
- tektite 253, 268
Terrestrial Planet Finder 190
thermophile 311, 332
Titan 354, 396
topoisomerase 309
TPF-C 201, 216
trans-Neptunian 132
tree of life 313
tropopause 124
troposphere 124
tubulin 226
Tycho Brahe 550
- Vega 367
Venus 203, 205
Vesta 91, 93–94
Viking 400
virus 223, 309
VLBI 182
VLT 188
VLTI 182
- water 111, 200, 228, 347, 354, 355, 408
 cycle 399
 on Mars 107, 110, 113, 115
Wood–Ljungdahl cycle 311
- zircon 83–84