The Astrobiology Primer v2.0

Co-Lead Editors
Shawn D. Domagal-Goldman and Katherine E. Wright

Chapter Editors
Shawn D. Domagal-Goldman (Co-Lead Editor, Co-Editor Chapter 1, and Author)1,2,*
Katherine E. Wright (Co-Lead Editor, Co-Editor Chapter 1, and Author)3,4,*
Katarzyna Adamala (Co-Editor Chapter 3 and Author)5
Leigh Arina de la Rubia (Editor Chapter 9 and Author)6
Jade Bond (Co-Editor Chapter 3 and Author)7
Lewis R. Dartnell (Co-Editor Chapter 7 and Author)8
Aaron D. Goldman (Editor Chapter 2 and Author)9
Kennda Lynch (Co-Editor Chapter 5 and Author)10
Marie-Eve Naud (Co-Editor Chapter 7 and Author)11
Ivan G. Paulino-Lima (Editor Chapter 8 and Author)12,13
Kelsi Singer (Co-Editor Chapter 5, Editor Chapter 6, and Author)14
Marina Walter-Antonio (Editor Chapter 4 and Author)15

Authors
Ximena C. Abrevaya,16 Rika Anderson,17 Giada Arney,18 Dimitra Atri,13 Armando Azúa-Bustos,13,19
Jeff S. Bowman,20 William J. Brazelton;21 Gregory A. Brennecka,22 Regina Carns,23 Aditya Chopra,24
Jesse Colangelo-Lillis,25 Christopher J. Crockett,26 Julia DeMarines,13 Elizabeth A. Frank,27 Carie Frantz,28
Eduardo de la Fuente,29 Douglas Galante,30 Jennifer Glass31 Damhnait Gleeson,32 Christopher R. Glein,33
Colin Goldblatt,34 Rachel Horak,35 Lev Horodyskyj,36 Betül Kaçar,37 Akos Kereszturi,38
Emily Knowles,39 Paul Mayeur,40 Shawn McGlynn,41 Yamila Miguel,42 Michelle Montgomery,43
Catherine Neish,44 Lena Noack,45 Sarah Rugheimer,46,47 Eva E. Stüeken,48,49
Paulina Tamez-Hidalgo,50 Sara Imari Walker,13,51 and Teresa Wong52

*These two authors contributed equally to the work.
1NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
2Virtual Planetary Laboratory, Seattle, Washington, USA.
3University of Colorado at Boulder, Colorado, USA.
4Present address: UK Space Agency, UK.
5Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, Minnesota, USA.
6Tennessee State University, Nashville, Tennessee, USA.
7Department of Physics, University of New South Wales, Sydney, Australia.
8University of Westminster, London, UK.
9Oberlin College, Oberlin, Ohio, USA.
10Division of Biological Sciences, University of Montana, Missoula, Montana, USA.
11Institute for research on exoplanets (iREx), Université de Montréal, Montréal, Canada.
12Universities Space Research Association, Mountain View, California, USA.
13Blue Marble Space Institute of Science, Seattle, Washington, USA.
14Southwest Research Institute, Boulder, Colorado, USA.
15Mayo Clinic, Rochester, Minnesota, USA.
16Instituto de Astronomía y Física del Espacio (IAFE), UBA—CONICET, Ciudad Autónoma de Buenos Aires, Argentina.
17Department of Biology, Carleton College, Northfield, Minnesota, USA.
18University of Washington Astronomy Department and Astrobiology Program, Seattle, Washington, USA.
19Centro de Investigación Biómédica, Universidad Autónoma de Chile, Santiago, Chile.
20Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, USA.
21Department of Biology, University of Utah, Salt Lake City, Utah, USA.
22Institut für Planetologie, University of Münster, Münster, Germany.
23Polar Science Center, Applied Physics Laboratory, University of Washington, Seattle, Washington, USA.
24Planetary Science Institute, Research School of Earth Sciences, Research School of Astronomy and Astrophysics, The Australian National University, Canberra, Australia.
Chapter 1. Introduction—What Is Astrobiology?

1.1. What is astrobiology?

Astrobiology is the science that seeks to understand the story of life in our universe. Astrobiology includes investigation of the conditions that are necessary for life to emerge and flourish, the origin of life, the ways that life has evolved and adapted to the wide range of environmental conditions here on Earth, the search for life beyond Earth, the habitability of extraterrestrial environments, and consideration of the future of life here on Earth and elsewhere. It therefore requires knowledge of physics, chemistry, biology, and many more specialized scientific areas including astronomy, geology, planetary science, microbiology, atmospheric science, and oceanography.

However, astrobiology is more than just a collection of different disciplines. In seeking to understand the full story of life in the Universe in a holistic way, astrobiology asks questions that transcend all these individual scientific subjects.

Astrobiological research potentially has much broader consequences than simply scientific discovery, as it includes questions that have been of great interest to human beings for millennia (e.g., are we alone?) and raises issues that could affect the way the human race views and conducts itself as a species (e.g., what are our ethical responsibilities to any life discovered beyond Earth?).

1.2. Have we already found life beyond Earth?

No. There have been many exciting discoveries that suggest life is possible on other planets and moons, but we have not yet detected any definite signs of life beyond Earth. That does not necessarily mean life exists only on Earth, but
only that there is not yet compelling evidence for its existence elsewhere. Space missions have explored only a tiny portion of our solar system, and in the few years since we first discovered planets around other stars, the number of such exoplanets known has increased into the hundreds. The search for life beyond Earth has therefore only just begun.

Astrobiology is an exciting subject with new ideas that can easily capture the imaginations of both scientists and nonscientists. But like all areas of science, new ideas are subject to detailed scrutiny by the scientific community as part of the quality control process. Only ideas that are considered to be well-supported by evidence are accepted by the scientific community as a whole. This can lead to lively debate within the scientific community and sometimes within the public arena. Whether the new idea is ultimately accepted or not, the process of testing the evidence results in increased knowledge and understanding, not just about the possibility for life beyond Earth but about our own planet as well. Two high-profile announcements within the field of astrobiology illustrate this.

In 1996, scientists announced that they had found evidence of fossilized life contained within martian meteorite ALH84001 (McKay et al., 1996). In subsequent years, astrobiologists completed research on the meteorite itself and on the types of “biosignatures” contained within it. As a result, we now know that abiological processes could have created these signals, and the meteorite is not therefore widely considered proof of life beyond Earth (Bradley et al., 1996, 1997; Frankel and Buseck, 2000; Buseck et al., 2001; McKay et al., 2003; Treiman, 2003). For example, one of the “biosignatures” cited by McKay et al. were structures in the meteorite that appeared to resemble fossilized bacteria when viewed under an electron microscope, but later work showed that similar structures can be produced as an artifact of the techniques used to prepare mineral samples for electron microscopy (Bradley et al., 1997). Although the claims made by McKay et al. have not been generally accepted, in the process of testing them astrobiologists advanced research on a variety of topics, including the minimum size of an individual cell and production mechanisms for tiny grains of magnetite. Additionally, astrobiologists were forced to reexamine what constitutes conclusive evidence for past or present life.

More recently, a team announced the discovery of a bacterium they claimed could substitute arsenic for phosphorus in its DNA (Wolfe-Simon et al., 2011). These claims have been largely rejected by the scientific community, which has generally concluded that they are not proven by the evidence presented in the paper (Benner, 2011; Borhani, 2011; Cotner and Hall, 2011; Csabai and Szathmáry, 2011; Foster, 2011; Oehler, 2011; Redfield, 2011; Schoepp-Cothenet et al., 2011; Erb et al., 2012; Reaves et al., 2012). Nevertheless, this paper has stimulated a very active debate, and further research will undoubtedly lead to improved understanding of microbial arsenic metabolism. For example, investigations into the claims made by the paper have already led to the hypothesis that the bacteria isolated by Wolfe-Simon et al. may contain a high-affinity phosphorus transporter that is stimulated by arsenic (Foster, 2011).

1.3. What is the Astrobiology Primer?

The Astrobiology Primer is designed to provide a basic, but comprehensive, introduction to the field of astrobiology. It is longer than a typical review paper but much shorter in length than a textbook, with the goal of being detailed enough to provide a brief overview of the variety of questions investigated by astrobiologists.

The Astrobiology Primer is the product of a strong, vibrant, early-career astrobiology community. This is the second version of the Primer, and like the first (Mix et al., 2006), it is a grassroots effort, written by graduate students and postdoctoral researchers. In total, we are a group of 49 authors and editors from 14 different countries. This second edition has been rewritten from scratch. It updates content and is organized around the questions that currently drive research in the field.

Our target readers for this document are other early-career astrobiologists, in particular graduate students who are new to the field, but we hope that it will also be useful to a wide range of people for both personal study and teaching.

This primer begins with the question of the nature of life (Chapter 2), then goes on to discuss the origins of life and its planetary environment (Chapter 3), the interactions of life with our planet through time (Chapter 4), what we know of habitability from these interactions (Chapter 5), what we know about the habitability of environments outside Earth (Chapter 6), and how we might search for life in those environments (Chapter 7). We close with chapters on the implications of this research for society (Chapter 8) and resources for astrobiologists (Chapter 9).

Chapter 2. What Is Life?

This simple question is surprisingly difficult to answer yet is fundamental to the success of astrobiological research. Imagine the difficulty of identifying life on other worlds without a clear understanding of what similarities it may or may not share with life on Earth. Or consider the limitations of interpreting possible origin-of-life scenarios without distinguishing between primitive life-forms and the nonliving entities from which they must have emerged. So far, and not without significant effort, no single definition of life has achieved universal acceptance. But the very exercise of attempting to define life reveals and tests its most essential characteristics.

2.1. Can we define life?

The goal is not to define “life,” the word, as it is used and understood in language, but rather to understand “life” as an objective concept that can guide scientific research (Oliver and Perry, 2006). Successful scientific definitions of life attempt to include everything that we already intuitively consider alive and exclude everything that we would not consider alive. For our purposes, such definitions should also be sufficiently broad to include unknown forms of life that independently arose on Earth or elsewhere. If these criteria are satisfied, the definition will be a useful guide in the search for life on other worlds and the study of its origin here on Earth (Ruiz-Mirazo et al., 2004; Oliver and Perry, 2006).

This objective, however, is made more difficult to achieve by our lack of a second known instance of life. Because all known life evolved from a single ancestor or ancestral community, every life-form shares a common set of inherited properties (Woese, 1998; Becerra et al., 2007; Theobald,
Consequently, it can be difficult to distinguish general features of life in the Universe from those specific to our own form of life (Gayon, 2010). Cleland and Chyba (2002) were the first to argue that, without additional examples of life and the greater understanding they may provide, it is even impossible to know whether the concept of life describes an objective natural phenomenon or a subjective category that cannot be perfectly defined (Cleland and Chyba, 2002; Robus et al., 2009).

Such objective phenomena, which philosophers call “natural kinds,” can be defined completely by principles of the natural world without depending on human-made conventions, for example water defined as the molecule H₂O. Cleland and Chyba used the term “bachelor,” defined as “an unmarried adult male,” to illustrate the sort of category that is not a natural kind. This definition “explain(s) the meaning of (the term) by relating (it) to expressions that we already understand,” but the term “bachelor” is not a natural kind in that the terms “adult” and “unmarried” are understood only as cultural concepts, not natural ones. Cleland and Chyba pointed out that “water” was a concept much like this because it became possible to describe water as its molecular formula. Life may be like water, a natural kind waiting for a scientific definition, or may be more like bachelor, a category that can be easily understood but that cannot be defined completely by natural principles.

A successful definition of life must not only delineate life and nonlife but also deal with the intermediate stages that may exist between life and nonlife. The origin of life presents just such a test. Our current understanding of life’s origin suggests that there were intermediate states through which all forms of life must emerge (Fry, 1995; Luisi, 1998; Perry and Kolb, 2004). In addition, the biosphere today includes entities that may represent intermediate states between life and nonlife. For example, viruses, which some do not consider bona fide organisms, possess many features similar to organisms and may present further evidence that there is a continuum between life and nonlife. All the arguments listed above demonstrate that the goal of creating a clear and objective scientific definition of life is not at all straightforward.

2.2. What are the common characteristics of life on Earth?

While life on Earth represents only one example, it is the only known example and, therefore, a good place to begin. Any universal characteristic of life on Earth may be universal either because it was inherited from a common origin or because it is a necessary feature of all life in the Universe. The lack of a second example of life frustrates our ability to conclusively differentiate between these two possibilities.

The chemistry of life is predominated by only a handful of carbon-based macromolecules common to all organisms: cellular membranes and intracellular compartments are primarily composed of a type of molecule called a “phospholipid,” a lipid with a charged phosphate group on one end; genetic information is stored and processed by the nucleic acids DNA and RNA; the catalytic and structural functions of the cell are performed mainly by proteins. Many of these so-called macromolecules are formed through the polymerization of subunits, for example, nucleotides (forming nucleic acids) and amino acids (forming proteins). The cellular pathway that creates proteins from genetic information is also common across life, and the genetic code that translates genetic information into protein molecules is also nearly universal (Knight et al., 2001).

In addition to a common biochemistry, all known life-forms exhibit many of the same general traits. Campbell and Reece (2002) listed the following traits that are common to life on Earth:

- **Ordered structure** refers to the high level of organization observed both within cells and within multicellular organisms as well as the bilateral or radial symmetry observed in many organisms.
- **Reproduction** can refer to either the nearly exact duplication of an organism or the production of a new organism through sex between two parent organisms.
- **Growth and development** refers to the processes by which organisms reach maturity, which can take drastically different forms depending on the type of organism.
- **Energy utilization** refers to the capture of energy from sources such as sunlight, inorganic chemical reactions, or organic material produced by other organisms, and subsequent use of that energy for cellular processes and the biosynthesis of cellular components.
- **Response to the environment** refers to the ability of organisms to sense external stimuli and alter their internal environments accordingly.
- **Homeostasis** refers to the maintenance of a steady internal environment regardless of the external environment.
- **Evolutionary adaptation** refers to the process by which populations of organisms adapt to external pressures such as environmental changes through evolution by natural selection.

Some phenomena that are clearly not alive demonstrate one or more, but not all, of these features. Crystals, fire, and hurricanes are all able to grow, reproduce, and utilize forms of potential energy. Crystals also demonstrate ordered structure, albeit composed of fewer distinct subunits. Elements of culture, or “memes” as originally coined by Dawkins (1976), reproduce and evolve by selection in ways that are similar to genes (Dawkins, 1976; Dennett, 1995). Many artificial intelligence algorithms are also based on learning by a process similar to evolution by selection (Dennett, 1995).

A much more ambiguous example is that of viruses, which are compartmentalized biological elements that contain genomes and reproduce by co-opting the metabolism of an infected cell. Viruses evolve, demonstrate ordered structure, and in some cases undergo a maturation process that could be construed as growth and development. But they cannot reproduce without co-opting the metabolism of their infected hosts. Viruses, themselves, also do not demonstrate homeostasis, energy utilization, or response to the environment. However, many pathogenic bacteria are also unable to reproduce outside a host cell. If complete metabolic autonomy were a requirement for life, animals, plants, and many other forms of life on Earth would not qualify.
Viruses are an especially important case for consideration because they are not only similar to cellular organisms in character, but they also use the same chemistry and genetic code that is discussed below. This similarity is necessary because, otherwise, viruses would not be able to take advantage of the host’s biosynthetic machinery to reproduce. Several recently discovered viruses have genomes and physical dimensions comparable in size to some cellular organisms and can be infected, themselves, by other viruses (La Scola et al., 2003, 2008).

Biological entities smaller than viruses are also capable of evolution by natural selection. For example, transposons are stretches of DNA found in genomes that, at a minimum, encode genes responsible for extracting or copying themselves and inserting themselves or their copies elsewhere in the genome. About 40% of the human genome is composed of transposons (Lander et al., 2001). Such minimal biological agents appear to blur the line between life and nonlife. The line becomes a bit less blurry through the observation that the majority of natural selection in the biosphere acts on genes rather than organisms or populations (Williams, 1966).

2.3. So what is life?

At a workshop in 2003, every member of the International Society for the Study of the Origin of Life was asked to give his or her definition for life, resulting in 78 different answers (Palyi et al., 2002). Most often, definitions of life draw on one or more of the attributes of life on Earth that were listed above, usually involving some combination of metabolism, reproduction, and evolution. For the reasons described in Section 2.2, no single feature of life is sufficient to define it, as each of these features is shared by other entities that are not generally considered life.

In his 1943 lecture series “What is Life?,” published in 1944, Erwin Schrodinger introduced several important concepts that would shape future considerations of the question. The first of these is the concept of genetic information or what he called “code-script.” At the time of this insight, DNA had not yet been confirmed as the genetic molecule, and its mechanisms of heredity and genetic encoding would not be described until the following decade. A second insight is that life appears to disobey the second law of thermodynamics. That is, the ordered structures of life do not entropically decay toward chemical equilibrium. Life is able to achieve this far-from-equilibrium state by “feeding” on disequilibrium in the environment. We now understand that this maintenance of order is achieved through the coupling of energy metabolism to biosynthetic metabolism.

Currently, the most popular definition of life was conceived by a NASA effort to create a “working definition” for their Exobiology and Astrobiology research programs. The result, most often cited as Joyce (1994), defines life as “a self-sustaining chemical system capable of Darwinian evolution.” A major strength of this simple and concise definition is that it distinguishes life by the evolutionary process rather than its chemical composition. The definition alludes to biological chemistry through the general term of “self-sustainment.” By invoking evolution as a foundational feature of life, the definition also separates the property of life from the characteristics of individual organisms, which are not, themselves, capable of Darwinian evolution. This is because the evolutionary forces discovered by Darwin (e.g., natural selection) operate through the reproductive success of organisms rather than the alteration of individual organisms during their own lifetimes.

Critics of the definition find that the stipulation of Darwinian evolution is overly specific (Cleland and Chyba, 2002). They argue that an extraterrestrial life-form may have a genetic system that evolves through changes accumulated during an organism’s lifetime (i.e., Lamarckian evolution) rather than Darwinian selection processes. Nevertheless, the NASA/Joyce definition is perhaps the most widely accepted definition of life within the astrobiological research community.

2.4. How should we use our concepts of life?

Even a less-than-perfect definition of life may be employed in the search for life on other worlds. For example, any “self-sustained chemical system” must differ in composition from its surroundings. Because life must also be “capable of undergoing Darwinian evolution,” this will result in successful species that amplify this chemical signature. McKay (2004) proposed that this chemical signature is a universal feature of all life and may be used for life detection. But while the Joyce (1994) definition gives us a useful way to identify candidate life-forms, it also requires that the life-form be verified by no less than clear evidence of evolution (Luisi, 1998; Cleland and Chyba, 2002).

A more pragmatic approach, and the only one used for life-detection experiments so far, is to look for signatures of life that are similar to those that would be produced by life on Earth (see Chapter 7). The first such attempts were made by experimental modules on NASA’s Viking Mars landers in the mid-1970s (Klein et al., 1976). These attempts focused on the search for the organic molecules of which Earth-based organisms are composed, and experiments designed to detect the production and consumption of gases by any organisms using similar metabolisms to those used by Earth-based organisms. A more recent approach to develop a life-detection experiment is the Life Marker Chip, designed to detect a broad range of biomolecules in either their pristine or degraded states (Parnell et al., 2007). To do so, the Life Marker Chip would consist of an array of antibody proteins that are designed to bind biomolecules common to life on Earth.

Scientific research will continue to enhance our understanding of life on Earth and its effect on the planet. Considering which of the traits demonstrated by life on Earth may be universal to all life in the Universe will expand our understanding even further. With this broader concept of life, future astrobiological searches should be capable of detecting a greater variety of life-forms. If future astrobiological missions are successful in finding life beyond Earth, they may fundamentally change our concept of life. In this way, the effort to find life on other worlds and the effort to define life universally will always be interconnected.

Chapter 3. How Did Earth and Its Biosphere Originate?

The study of how life began on Earth offers information we can use to consider the possibilities for the existence of life elsewhere. This chapter starts with the availability of
elements that could form the building blocks of life (Section 3.1). It continues with the origin of our planet and its volatile inventory (compounds with a low vaporization temperature/pressure) and considers what we know about the environment of early Earth (Section 3.2). The final section (Section 3.3) covers what we can determine about events that could have led to the origin and earliest evolution of life on Earth, given what we know about the environmental context.

3.1. How do elements, stars, and planets form?

3.1.1. How are the elements formed and distributed? All chemical elements are formed in nucleosynthetic processes involving addition or removal of nucleons (protons and neutrons). The earliest synthesis of elements in the Universe took place during Big Bang nucleosynthesis, which produced the vast majority of the hydrogen and helium present in the Universe today. Since then, stellar nucleosynthesis in stars has produced the majority of other naturally occurring elements (Burbidge et al., 1957) (Fig. 1).

Other known processes that have made minor contributions toward the production of elements in the Universe include photodisintegration of existing nuclei by gamma rays (cosmic ray spallation); radioactive decay of unstable elements; fission reactions in natural nuclear reactors; and recently, high-energy particle accelerators, nuclear bombs, and nuclear reactors. Together with Big Bang nucleosynthesis and stellar nucleosynthesis, these processes collectively yield all the isotopes of elements that exist in nature.

Big Bang nucleosynthesis began approximately 13.8 billion years ago, during the first 3 min after the Big Bang. The high temperature and density of the Universe during this time resulted in energies and collision frequencies that were high enough to overcome the electrostatic repulsion between nuclei and bring nucleons close enough for the attractive nuclear force to be sufficiently strong to achieve fusion. A series of fusion reactions of different isotopes could then produce the heavier elements helium and lithium. This period lasted for about 17 min and was responsible for the formation of the stable isotopes of hydrogen, helium, and lithium (the three lightest elements in the periodic table) from combinations of fundamental particles (Weinberg, 1993). Unstable (radioactive) isotopes of hydrogen and beryllium were also synthesized but decayed or fused with other nuclei to form stable isotopes.

Over time, this matter (also referred to as baryonic matter) fell into existing cold matter clumps to form clouds mostly composed of neutral molecular hydrogen. As densities increased in regions within the cloud, gravitational collapse triggered cloud fragmentation. Further collapse of the clouds led to increasing pressures and temperatures at their cores. Eventually, temperatures and densities in these cores reached levels required for nuclear fusion. This is how the first stars were born, beginning the era of stellar nucleosynthesis of heavier elements (Brom, 2013). This era of element formation began about 100 million years after the Big Bang and continues to this day (Fig. 2).

The nucleosynthetic pathways that are possible in a star are closely coupled to stellar evolution; different types of stars synthesize a range of different elements during their lives.
lifetimes (Baross and Sullivan, 2007; Ryan and Norton, 2010). The successive nuclear fusion processes, commonly referred to as nuclear burning (not to be confused with “burning” meaning combustion in the presence of oxygen), are hydrogen burning, helium burning, carbon burning, neon burning, oxygen burning, and silicon burning. These nuclear burning (or fusion) reactions occur sequentially within a single star or series of different generations of stars. Collectively, they create isotopes as heavy and energetic as $^{56}$Ni through exothermic (energy-yielding) reactions that release the energy needed to sustain stars (see Section 3.1.2).

Isotopes with masses larger than $^{56}$Ni can only be generated with endothermic (energy-consuming) reactions. Therefore, elements heavier than Ni require energy to form and are not produced through the exothermic nuclear fusion reactions that take place at the centers of stars. (While $^{56}$Ni is the last fusion product produced in the cores of high-mass stars, this isotope undergoes radioactive decay into $^{56}$Fe, significantly increasing the abundance of iron relative to other first-row transition elements in the periodic table. This phenomenon is often referred to as the “iron peak.”)

Elements heavier than iron are assembled when there is a high abundance of highly energetic neutrons and protons in the star. Such scenarios often occur in the last few seconds to minutes before the end of the star’s life, as in core-collapse supernovae (Arnett, 1996) (see Section 3.1.2). There are a number of “bottlenecks” in various nucleosynthetic pathways due to the dependence of reaction rates on each other, and these determine the relative abundances of the isotopes of the elements that are created (Lodders, 2010).

Generations of stars have gradually converted the products of Big Bang nucleosynthesis into heavier elements such as oxygen, iron, silicon, and magnesium. These elements are now available to form planets, moons, and life (Fig. 3). Most rocks on the surface of Earth are composed of minerals made of lithophilic (“rock-forming”) elements (Fig. 4). Input of material from comets and primitive meteorites such as chondrites in Earth’s early history led to a biosphere (containing life and its environment) rich in the volatile elements hydrogen and oxygen (mostly as water) along with carbon, nitrogen, sulfur, and phosphorus (mostly as dissolved salts or gases) (Chyba et al., 1990).

3.1.2. How do stars form, evolve, and die? A star forms when a cloud of predominantly hydrogen gas collapses under its own gravity. As it collapses, the temperature and pressure at the core become great enough to fuse hydrogen into helium. Further collapse of the star is halted when the pressure in the star’s interior is high enough to counteract the force of gravity.

Nucleosynthesis is the process by which stars create progressively heavier elements (see Section 3.1.1). Starting at the core, the nucleosynthetic region within the star moves outward as its temperature within that region increases. As each region hosts nuclear fusion of progressively heavier elements at higher temperatures, different shells of nuclear burning define an onionlike structure within the star as illustrated in Fig. 5. The energy-releasing fusion process continues until iron is formed in the core of the star.
FIG. 3. Elemental composition of the Sun (a typical star) by mass. The word “metal” here is used in the way commonly applied to stellar compositions and includes all elements heavier than H and He. (Credit: Martin Asplund)

FIG. 4. The abundance of elements (by number of atoms normalized to silicon) in Earth’s crust. The crust is depleted in siderophilic (“iron-loving”) elements relative to the bulk Earth as siderophiles were partitioned into the Earth’s core. Compared with the mantle, Earth’s crust is enriched in lithophilic (rock-forming) elements. (Credit: Aditya Chopra)
Under the conditions found in stellar interiors, iron cannot be turned into a heavier element through the fusion process because, unlike fusion of lighter elements, iron fusion requires energy. Thus, once iron is formed at the core of a star, fusion ceases to operate in that region of the star. This places an upper limit on the amount of radiation pressure produced by fusion, and because radiation pressure is what prevents gravitational collapse, this leads to the death of stars. This same process also limits the heaviest element that can be formed in any given star. Iron is the heaviest element that can form in the most massive stars, but less massive stars may only be capable of forming lighter elements.

Lifetimes of stars are thought to be as long as trillions of years for the lightest stars and as short as a few millions of years for the most massive stars (Phillips, 1999). There are two pathways a dying star can take (Fig. 6). For low- to intermediate-mass stars, the core collapses to become a “white dwarf”—a hot, compact ball of helium, carbon, and oxygen—while the outer layers are ejected into space. In the most massive stars, the collapse initiates an explosion, known as a supernova, that triggers the formation of elements heavier than iron (see Section 3.1.1). Depending on the star’s initial mass, the remaining core collapses to become either a neutron star (a dense city-sized ball of neutrons) or a black hole (an entity whose gravity is so strong that even the fastest particles in the Universe, photons, cannot escape its gravity well; as a result, black holes emit no light). In most cases, the newly formed elements that are not incorporated into the neutron star or black hole are dispersed back into the interstellar medium and become incorporated into the next generation of stars and planets (Whittet, 2003).

3.1.3. How do protoplanetary disks and planetary systems form? Protoplanetary disks are diffuse regions of gas and dust and are the eventual birthplaces of stellar and planetary systems. During the gravitational collapse of a molecular hydrogen cloud, conservation of angular momentum causes the cloud’s rotation to increase, and the rotation causes the cloud to flatten into a disk. The center of the disk contains the newly forming star, and the disk is the site of planet formation.

Planetary bodies are widely believed to have formed through the method of core accretion (Ida and Lin, 2004). Under this paradigm, dust accretes over time to produce rocky bodies known as planetesimals that range from kilometer size to approximately the size of Earth’s Moon. These bodies interact gravitationally with each other, colliding and accreting to produce planetary embryos (Moon- to Mars-sized solid bodies) or being broken apart by violent...

FIG. 5. Onionlike structure caused by progressive nucleosynthesis in stars. Nearest to the core are the sites of the hottest and heaviest nuclear fusion, while regions farther away near the surface host the coolest and lightest nuclear fusion reactions. (Credit: NASA)
collisions. An alternative planet formation model that has been hypothesized is that of gravitational instability (Zel’Dovich, 1970; Boss, 1997). In the instability model, protoplanetary disks fragment during the collapse of the disk, creating high-density regions that rapidly pull in surrounding material to form planetary bodies. This model, however, has not been widely adopted by the modeling community and encounters difficulties in producing the observed architecture of extrasolar planetary systems. As such, we will focus on the accretion model.

In the planetary accretion model, planets with relatively high metallic and rocky content, such as Mercury and Earth, are expected to form closer to their host star. Conversely, planets with greater volatile content (molecules that have low boiling points), such as the giant planets of the outer Solar System, are expected to form farther from their host star. The latter group of planets can become much more massive than their rocky counterparts due to addition of extra mass in the form of accreted volatiles. This explanation is consistent with—and historically driven by—the arrangement of planets in the Solar System: smaller, rocky planets exist closer to the star and larger, volatile-rich planets farther from the star.

However, Jupiter-mass and Neptune-mass exoplanets have been found closer to their host than Mercury is to the Sun (Butler et al., 2006; Borucki et al., 2010), so planetary migration through the disk has been suggested to explain how these hot giant planets moved from their birth locations to their current locations (see below). Combined, core accretion and orbital migration form the general framework for planet formation, both within our solar system and within extrasolar planetary systems (Kley and Nelson, 2012).

Two dominant models have emerged under the planetary accretion paradigm to explain the formation of planets and long-term evolution of the Solar System: the Grand Tack model, which focuses on planet formation (Walsh et al., 2011), and the Nice model, which focuses on long-term evolution (Gomes et al., 2005; Morbidelli et al., 2005; Tsiganis et al., 2005; Levison et al., 2011). Both models involve large-scale migration of the giant planets before settling in their current orbits and can explain a variety of features of the Solar System. However, they focus on very different time periods (Nice occurred approximately 500 million years after the Solar System formed, while the Grand Tack occurred within the first 5 million years). The two models also have somewhat different end results in that the Nice model reproduces the outer Solar System well, while the Grand Tack reproduces the inner Solar System.

The Grand Tack model addresses formation events that occurred in the earliest stages of the formation of the planets within the Solar System. Under this model, Jupiter and subsequently Saturn began their formation considerably closer to the Sun than they are now (approximately 3.5 AU; Walsh et al., 2011). Once large enough, both protoplanets began to experience type II migration wherein they were coupled to the
gas disk and slowly migrated inward, along with the gas. This migration brought the bodies in to 1.5 astronomical units (AU), where Saturn effectively caught up with Jupiter (as Saturn-mass planets tend to migrate faster) and the cleared area around each of the bodies overlapped. This caused the planets' migration to change direction (i.e., 'tack,' hence the name of the model) and begin to migrate outward. As they migrated outward, they encountered Uranus and Neptune and caused the outward migration of these two planets to their current orbits.

The Grand Tack model has the benefit of explaining the structure and composition of the asteroid belt. There are two very different types of asteroids in the asteroid belt—ordinary chondrites and carbonaceous chondrites—which have very different compositions and therefore must have formed in different locations. According to the Grand Tack model, the giant planets scatter planetesimals throughout the disk as they migrate outward from 1.5 AU. Some of these are scattered inward, which results in the introduction of carbonaceous chondrites to the asteroid belt and explains the presence of two vastly different types of asteroids within the asteroid belt—ordinary chondrites in the inner asteroid belt and carbonaceous chondrites in the outer belt. The Grand Tack model allows for the combination of the two types in one location.

The Grand Tack model also has the added benefit of potentially resolving the longstanding issue of Mars' low mass. The formation of Mars has been difficult to simulate correctly, as most models produce planets that are too massive (We-therill, 1991; O'Brien et al., 2006; Raymond et al., 2009). One solution to this problem is to truncate the disk of planetesimals that accrete to produce the terrestrial planets at 1 AU (Hansen, 2009). The Grand Tack model provides a mechanism to achieve this, as the presence of (a still-forming) Jupiter at 1.5 AU results in a pileup of planetesimals at 1 AU and depletes the feeding zone for proto-Mars.

The Nice model, on the other hand, is focused on the evolution of the Solar System and results from the giant planets (Jupiter, Saturn, Neptune, and Uranus) initially being located farther from the Sun and closer to each other. Dynamical interactions with a large remnant planetesimal population result in orbital instabilities for Jupiter and Saturn. These orbital instabilities resulted in gravitational interactions with Uranus and Neptune, also affecting their orbits. Subsequent gravitational interactions between the giant planets and a planetesimal disk on the outer edges of the Solar System not only produce the influx of impactors observed in the inner Solar System during the Late Heavy Bombardment (LHB) period (see Section 3.2) (Gomes et al., 2005) but also serve to remove the majority of the planetesimal population from the outer Solar System, which explains the current dearth of trans-Neptunian bodies. The Nice model can also explain not only the structure of the asteroid belt but also some of the compositional variations observed therein (Levison et al., 2009), along with the structure of both Jupiter and Neptune's trojan population (Morbidelli et al., 2005; Tsiganis et al., 2005; Sheppard and Trujillo, 2010). This initial model has been modified, with the giant planets being suggested to have initially started in a quadruple resonance (Levison et al., 2011), migrating largely together due to gravitational interactions with the outer planetesimal disk.

Combined together, the Nice and Grand Tack models may represent the formation and evolution pathway of the planets within our solar system and potentially within other extrasolar systems. For further details regarding terrestrial planet formation models, see Morbidelli et al. (2012) for a Solar System–based discussion and Raymond et al. (2014) for discussion of planet formation within both the Solar System and extrasolar planetary systems. See also Chapter 6 for more information on our current knowledge of planets in other solar systems.

Once a planet forms (under any model) there can be further changes that result in the differentiation of elements between layers (Chambers, 2004). If enough heat is present in a planet’s interior, the planet effectively melts into a liquid form, so different elements can rise or sink depending on their density, causing differentiation between metal-rich (siderophile), silica-rich (lithophile), and/or gas-rich (atmospheric) components. During this differentiation, relatively dense components such as iron, nickel, or metallic hydrogen tend to sink to the center to form the planet’s core, whereas relatively less dense materials such as silica and gases become preferentially incorporated into the mantle and crustal layers above. Differentiation plays a major role in the distribution of elements in a planetary body.

3.2. How did Earth, including its oceans and atmosphere, form?

Earth formed via accretion (see Section 3.1.3) from the solar nebula 4.53–4.47 billion years ago (Ga). The age of Earth was established based on the combination of both Earth-based and meteoritic samples. Earth-based evidence includes the age of the oldest minerals currently persevered on Earth’s surface (4.4 Ga; Wilde et al., 2001) and the oldest whole rock known to exist on Earth (4.208 Ga; O’Neil et al., 2008). Meteoritic evidence includes the age of the first solids to condense within the Solar System, known as calcium-aluminum inclusions (4.565 Ga; Connelly et al., 2008) and radiometric dating of various meteorite samples (ages clustering from 4.68 to 4.50 Ga).

Earth accreted and differentiated over a 10–150 million-year time period, based on the $^{182}\text{Hf}$/$^{182}\text{W}$ isotope system. $^{182}\text{Hf}$ decays into $^{182}\text{W}$ with a half-life of 9 million years. Hafnium is a lithophile (i.e., an element that preferentially bonds with silicon) and as such partitioned into proto-Earth’s silicate-rich crust. Tungsten, on the other hand, is a siderophile and as such partitioned into the core during Earth’s formation. Thus, the observed $^{182}\text{W}$ enrichment in Earth’s crust (relative to other W and Hf isotopes) must be a result of $^{182}\text{Hf}$ decay, meaning formation and differentiation must have occurred while $^{182}\text{Hf}$ was still active and present within the system. Thus, the formation of Earth commenced no earlier than 10 million years after the formation of the Solar System (Kleine et al., 2004; Jacobsen, 2005; Touchol et al., 2007). There is some debate about whether a rapid accretion (10–30 million years) or slower accretion (50–150 million years) occurred, based on the timing of the Moon-forming impact (discussed below). This issue currently remains unresolved. The $^{182}\text{Hf}$/$^{182}\text{W}$ isotope system has also been applied to dating Mars (Dauphas and Pourmand, 2011).

The Hadean marks the first geological eon in Earth’s history and spans the period from the end of the accretion to
the beginning of the Archean eon at 3.8 Ga. During the Hadean, impactors pummeled proto-Earth. The most dramatic event during this period occurred within the first 100 million years of Earth’s history. According to the most widely accepted hypothesis, this is when young Earth collided with a Mars-sized object, which led to the formation of the Moon (Canup and Asphaug, 2001). Heat from the impact left Earth’s surface in a molten state, and due to the lack of an atmosphere immediately following impact, cooling occurred quickly. Within 150 million years a solid basaltic crust reformed. Steam escaping from the crust and gases released from volcanoes generated the prebiotic atmosphere and contributed to the formation of oceans (see Section 3.2.1). The bombardment of Earth by impactors continued after the Moon-forming impact, resulting in the delivery of the “late veneer.” This late veneer of material is evidenced by the presence of highly siderophilic elements within Earth’s mantle and crust (as opposed to the core, where such elements would preferentially appear). Eventually, delivery of the late veneer material essentially ceased. According to the most well-known, though not universally accepted, hypothesis, this terminated with the LHB period (between 4.1 and 3.8 Ga), which marks the most intense period of bombardment after Earth’s formation. Some geological models suggest that Earth could not have been continuously habitable during this period because the impacts would have sterilized the planet (Maher and Stevenson, 1988). However, there is some debate as to the global lethality of such impacts (Sleep et al., 1989), and more recent modeling suggests that, if life existed before the LHB, it could have survived (Abramov and Mojsis, 2009), which leaves an absolute time frame for life’s origin undetermined.

3.2.1. How were volatiles brought to Earth? The emergence of life (as we know it) requires the presence of a variety of volatile elements and species. However, Earth’s feeding zone is believed to have been dry and depleted in volatiles (species that readily enter into the gaseous phase due to a lower temperature of vaporization), especially water (based on the volatile-poor composition of the innermost asteroids). As such, delivery of these species (or their precursors) needed to occur during the formation and early evolution of Earth. Water delivery is understood to have occurred throughout the formation of Earth, primarily through the accretion of asteroidal material, with comets and other sources providing limited water. Organics were delivered via asteroids, with a very small amount being derived from cometary material.

Given their ice-rich composition, comets were originally suggested as the primary source of Earth’s water (e.g., Owen and Bar-Nun, 1995; Delsemme, 2000). This idea has since been largely discredited due to the low probability of cometary collisions with Earth (Morbildelli et al., 2000) and the high D/H ratio of many, though not all, comets (e.g., Marty and Yokochi, 2006). Although water-poor in comparison with comets, asteroidal material has also been suggested as a water source, given its higher probability of colliding with Earth (Morbildelli et al., 2000; Raymond et al., 2004). This scenario has been altered somewhat by the Grand Tack model (discussed previously, see Section 3.1.3). Currently, our understanding is that the bulk of Earth’s current water was derived from material analogous to asteroids and planetary embryos, delivered early in Earth’s history, and later from larger planetary embryos that formed the outer, volatile-rich regions of the protoplanetary disk. This model has traditionally been difficult to simulate (Raymond et al., 2009), yet the Grand Tack model has proved to be promising, with the migration of material inward and then later outward within the Solar System providing the required mechanism to deliver volatile-rich material to a dry proto-Earth. Although it is unlikely that a significant fraction of the earliest water delivered would be retained, the later water delivery by embryos would supply sufficient water to account for Earth’s water budget. No more than 10% of Earth’s water was delivered by comets originating from the trans-uranian region, based on the differences in observed D/H ratio between comets and Earth’s current water supply (Morbildelli et al., 2000).

Impacts from comets and asteroids delivered a variety of organics to early Earth, even though they provided only a very small fraction of Earth’s water budget (Morbildelli et al., 2000). The interstellar medium, asteroids, and meteorites have been shown to host organics necessary for life (Sephton, 2005; Ehrenfreund and Cami, 2010; Pizzarello and Shock, 2010). The inventory of organics synthesized from the elements in interstellar space contains at least 120 identified species, including the simplest amino acid, glycine (C₂H₅NO₂). Additionally, formaldehyde, cyanide, acetaldehyde, water, and ammonia—all precursors in different synthesis pathways for amino acids—have been detected. (Amino acids are molecules that contain an amine group, a carboxylic acid group, and a side chain, which play a prominent role in modern biology as the primary building blocks of proteins.) During the formation of the Solar System, some of this interstellar organic material would have been integrated in the presolar nebula (the gas cloud from which the Solar System condensed—see Section 3.1) and incorporated into comets, asteroids, and other planetesimals.

Impacts of these objects (specifically asteroids) delivered volatile organics to ancient Earth (Chyba and Sagan, 1992). Sample return missions provide some of the best data about the composition of early impactors, as returned samples will have the most pristine records of the earliest Solar System (Matzel et al., 2010; Nittler, 2010). Additionally, comets likely contributed small amounts of carbonaceous material during the heavy bombardment (4.5–4 Ga). A surprise from the initial analysis of the Stardust samples retrieved from the comet Wild 2 was that the mineralogy was similar in composition to inner Solar System asteroids (Ishii et al., 2008). Other major extraterrestrial sources of carbon include interplanetary dust particles and carbonaceous meteorites (Chyba et al., 1990; Pizzarello and Shock, 2010). The Japan Aerospace Exploration Agency (JAXA) mission Hayabusa returned samples from the asteroid Itokawa, the study of which confirmed that asteroids are also similar in composition to ordinary meteoritic chondrites (Tomoki et al., 2011), and the OSIRIS-REx mission will return samples from a near-Earth asteroid. These extraterrestrial sources would have added to the endogenous inventory of carbon already existing on early Earth (Chyba and Sagan, 1992). Phosphorus is often the limiting resource for biological systems on Earth, as it is essential for all known life, but the geochemical cycling of phosphate is slow, and most phosphate minerals are insoluble (Pasek, 2008). A challenge for prebiotic chemistry is, therefore, to find sources for reactive and soluble phosphorus (Bryant and Kee, 2006; Pasek, 2008). In particular, meteorites may have been an important source of reactive phosphorus—such as the P-bearing ions pyrophosphate and...
triphosphate, key components for metabolism in modern life—
a likely requirement for the emergence of life (Pasek and
Laureretta, 2005).

Once delivered to proto-Earth and early Earth, geological
activity provided many of the volatiles necessary for the
emergence of life. In particular, Earth’s early atmosphere
formed via volcanic degassing (see Section 3.2.2), which
would have provided \( \text{N}_2 \), \( \text{CO}_2 \), and water, with small quan-
tities of \( \text{H}_2 \), \( \text{CO} \), and \( \text{CH}_4 \), and negligible amounts of \( \text{O}_2 \)
(Zahnle et al., 2010). This composition was at least mildly
reducing, although how reducing it was is debated (Tian et al.,
2005; Catling, 2006). Reactions of these volatiles in a weakly
reducing early atmosphere could have produced molecular
precursors to life (see Section 3.3). In addition, recycling of
bioessential elements such C, N, O, and P on modern Earth is
important to sustained habitability (see Chapter 5) and may
have been crucial in the early development of life on Earth
(Sleep, 2010).

3.2.2. What was the environment of prebiotic Earth like?
There is evidence that life on Earth was well established as
early as 3.5 Ga (Schopf, 1993; Wacey et al., 2011) and
possibly as early as 3.8 Ga (Mojzsis et al., 1996). However,
the ages of the most ancient fossils are difficult to determine
and typically subject to debate (e.g., Brasier et al., 2002;
Bradley et al., 2009). If life was in fact thriving as early as
3.8 Ga or even 3.5 Ga, two possible scenarios emerge. The
first is that life may have originated before or during the
LHB and survived this tumultuous period in the early his-
tory of Earth. This viewpoint is strengthened by evidence
that indicates Earth may never have been fully sterilized
during the LHB (Abramov and Mojzsis, 2009). A second
possibility is that, just after the LHB, the conditions on
primitive Earth were conducive to the rapid emergence of
life (Oberbeck and Fogelmann, 1989; Sleep et al., 2001).

Over time, as the He/H content of the Sun’s core increases,
it burns increasing amounts of He relative to H in stellar
fusion reactions (see Section 3.1.2 for a description of this
process). Because He burning gives off more energy than H
burning, the Sun has become brighter over the course of
Earth’s geological history. Going back in time, the Sun was
about 30% less luminous 4.5 billion years ago than it is today
(Sagan and Mullen, 1972) and 20% less luminous 2.5 billion
years ago than it is today. This leads to the faint young Sun
paradox: a planet with Earth’s current atmospheric compos-
tion orbiting the faint early Sun would have had cold global
temperatures that would have frozen all water at the surface of
the planet, yet there is evidence for liquid water on early
Earth. One resolution is that the early atmosphere was prob-
ably very different from the atmosphere of modern Earth and
may have contained higher levels of greenhouse gases (Pav-
lova et al., 2000; Haqq-Misra et al., 2008). This is still an area
of extensive research, and no definite theory has been widely
accepted.

By the start of the Archean, global oceans covered most
of the surface of Earth. Evidence of mineral interactions
with water suggest that oceans may have started to form by
4.2 Ga (Cavosie et al., 2005) or possibly as early as 4.4 Ga
(Wilde et al., 2001). Water would have degassed along with
volatiles such as \( \text{N}_2 \) and \( \text{CO}_2 \) to form the primitive atmos-
phere. Condensation of the steam produced from volcanic
degassing would have occurred as Earth cooled and led to
the formation of global oceans. However, it is widely agreed
that the primary source of water on Earth was delivery by
material similar to C-type asteroids (Morbidelli et al., 2000)
via the Grand Tack model, as discussed above.

3.3. How did life emerge?

Generally, five major stages of the origin of life can be
distinguished, as follows:

- First, the simplest organic building blocks were syn-
thetized, starting from the inorganic substrates.
- Second, small organic molecules started self-organizing,
giving rise to the larger, more complex organic molecules.
- Third, complex molecules started interacting with each
other, giving rise to the first reaction pathways, and the
network of reactions among molecules in different re-
action cycles began.
- Fourth, macrostructures arose from the spontaneous ag-
gregation of the simpler molecules, which led to com-
partmentalization, and the chemical reaction cycles
started to interact with the compartments.
- Fifth, the system was capable of being shifted from
chemical equilibrium, maintaining homeostasis and
undergoing evolution.

These stages are shown in Fig. 7. Events of the five de-
scribed stages could have been overlapping in time and
space, and it is often difficult, if not impossible, to uniquely
assign particular processes to one stage.

To understand how life could have originated, we need to
understand the types of environments that were present on
Earth, the energy sources and raw materials that were avail-
able, the chemical reactions that might have been possible
given the environmental conditions, and the processes that
would have set a sequence of events in motion that resulted in
the origin of life.

3.3.1. What do we know about the environment in which
life originated? Section 3.2.2 covers what we know about
the overall environment of early Earth (in the Hadean age,
4.5–3.8 Ga), but it is important to consider the dynamics and
composition of localized environments (microenvironments)
that could have existed and that might have provided suit-
able settings for the origin of life. Each would have provided
unique sources of energy, specific raw materials, and dif-
ferent mechanisms to facilitate chemical reactions. Apart
from organic synthesis, three broad categories of reactions
would have been required, including concentration of in-
redients from a dilute medium, selection of biochemically
useful compounds, and polymerization of monomers. Sev-
eral possible microenvironments are listed in Table 1. The
various sources of energy, the availability of raw materials,
and the numerous types of chemical processes that could
have taken place on early Earth are discussed in the sections
below. Stüeken et al. (2013) provided a more detailed re-
view of environmental linkages and their relevance to pre-
biotic chemistry.

3.3.2. Which energy sources were available? Energy
would have been required for a variety of necessities for the
origin of life: the synthesis of important prebiotic precursors,
concentration of potentially reactive solutes, energetically
FIG. 7. General schemes of possible stages of prebiotic evolution. (Credit: K. Adamala, adapted from scenario proposed by Eigen and Schuster, 1982)
3.3.3. What raw materials were present on early Earth?

One of the most commonly used definitions of life is that it is a chemical reaction system capable of Darwinian evolution (see Chapter 2). The complex chemical reaction systems that exist today must have originated as a result of a long series of simpler, more primitive processes. The initial raw materials for these chemical reactions are the individual chemical elements, which react with each other and combine to form molecules. Section 3.1 (above) explains what is known about how the elements were formed and incorporated into planets, while Section 3.2.1 covers the way in which volatile compounds, which would have been lost during the accretion of Earth due to the intense heat, would have been replenished. While many molecules, including organic carbon molecules, can form abiotically (without the intervention of life), one of the characteristics of life is the presence of highly complex organic carbon compounds that are not known to form by any nonliving processes (see Chapter 7 for more detail). Theories about the origin of life must therefore explain how complex organic molecules could have formed (see Section 3.3.5 below).

3.3.4. What types of prebiotic chemical reactions could have occurred? The characteristics of the microenvironments on early Earth would have directly influenced the type of chemical processes—prebiotic chemical reactions—that could have taken place and led to the origin of life.

3.3.4.1. Reaction medium. The environment of prebiotic reactions on Earth is considered to be predominately aqueous (Nisbet and Sleep, 2001). Some processes are postulated to take place in the gas phase in the atmosphere (Dobson et al., 2000), but the majority of prebiotic reactions need to be compatible with prebiotic Earth ocean conditions.

The prebiotic ocean was radically different from the modern ocean (Derry and Jacobsen, 1990). As the atmosphere on early Earth was non-oxidizing (Catling and Claire, 2005), the ocean would not have been oxidized, so it is believed to have contained high levels of the reduced form of iron (Fe^{2+}), which may have played a significant role in prebiotic redox reactions. Some have suggested that, unlike the oceans of today, the early ocean was more alkaline, lower in Mg and Ca cations, and lower in NaCl (Kempe and Degens, 1985). Others have concluded that the early ocean was similar to those of the present, perhaps even mildly more acidic given the higher abundance of atmospheric CO₂ gas (Grotzinger and Kasting, 1993). Large gradients in pH and alkalinity

<table>
<thead>
<tr>
<th>Potential settings</th>
<th>Energy sources</th>
<th>Major sources of organic matter</th>
<th>Concentration mechanisms</th>
<th>Mechanisms enhancing organic reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface ocean/lakes/ponds</td>
<td>Lightning, UV radiation, redox gradient</td>
<td>Meteorite deposition, Miller-Urey-type processes</td>
<td>Mineral surfaces, evaporation/rehydration</td>
<td>Clays at the bottom of lakes</td>
</tr>
<tr>
<td>Sea ice</td>
<td>Lightening, UV radiation, redox gradient</td>
<td>Uptake from seawater, atmospheric deposition</td>
<td>Brine pockets</td>
<td>Dehydration</td>
</tr>
<tr>
<td>Hydrothermal vents</td>
<td>Redox gradients, e.g., between H₂ and CO₂</td>
<td>Serpentization in peridotite-hosted vents</td>
<td>Carbonate or sulfide mineral pores</td>
<td>Metal sulfides, mineral surfaces, heat</td>
</tr>
<tr>
<td>Beaches</td>
<td>Lightning, redox gradient, radioactive radiation from accumulated uranium grains, UV radiation</td>
<td>Meteorite input, Urey-Miller-type processes, in situ production by nuclear fission</td>
<td>Evaporation/rehydration of pore spaces</td>
<td>Dehydration, nuclear fission by radioactive uranium</td>
</tr>
</tbody>
</table>

This table is not meant to be exhaustive but rather to list commonly discussed settings and requirements. All these environments, as well as many others, could have been interconnected by various mixing processes.
probably occurred in the vicinity of hydrothermal vents (Holm, 1992). On modern Earth, vent fluids can range from pH 2 to pH 11, depending on whether they are driven by magmatic heating or serpentinization, and it is likely that both processes occurred under prebiotic conditions.

Many prebiotic chemistry processes may have occurred in lagoons or pools, where tides drove cyclic changes of concentrations that provided the increase of concentration necessary for increasing the yields and efficiency of the prebiotic reactions (Nelson et al., 2001). Arguments against this theory are that the initial concentration of substrates required for the “lagoon theory” was higher than the possible concentration of organic compounds in the primordial ocean and that side reactions and the presence of inhibitors would have decreased the overall yield when the solution was concentrated (Shapiro, 2002). Nevertheless, the so-called “lagoon theory” has strong proponents (Orgel, 2004) and may have been an important component of the global chemical reactor (Stu¨eken et al., 2013). Another setting that has been gaining support in recent years is the oceanic crust with its associated hydrothermal alteration zones. Multiple experimental and theoretical studies have demonstrated that organic synthesis, polymerization, and selection of compounds can occur under hydrothermal conditions (Baross and Hoffman, 1985; Martin et al., 2008; He et al., 2010; Amend et al., 2013; Barge et al., 2014; Kreysing et al., 2015). The advantages of subaerial lagoons and hydrothermal systems were perhaps combined in terrestrial hot springs (Mulkidjanian et al., 2012; Schrenk et al., 2013). Importantly, hydrothermal vents may exist on icy moons such as Enceladus (Glein et al., 2015), which lacks continental lagoons. If life exists on those bodies, then it would underscore the importance of hydrothermal processes in prebiotic chemistry.

3.3.4.2. Temperature. The range of temperatures available for prebiotic reactions vary greatly. The building blocks of life can be synthesized in interstellar clouds and on comets, in temperatures approaching absolute zero, and near hydrothermal vents where temperatures may be over 100°C. Also, the temperatures postulated for the actual origin of life vary, some hypothesized favoring a hot environment (Miller and Lazcano, 1995), others favoring a cold environment (Bada and Lazcano, 2002). The average temperature on prebiotic Earth may have been higher than it is today (Arrhenius, 1887), but this is a topic of current debate. Given the lower luminosity of the young Sun and uncertainties about the abundance of greenhouse gases and continental area at the time, it is also possible that Earth cooled down relatively quickly to lower temperatures after its formation (Sleep et al., 2001; Rosing et al., 2010; Pope et al., 2012). For prebiotic chemistry, the average surface temperature is less important than the occurrence of liquid water at the surface. Evidence for liquid water dates back to 4.3 Ga (Mojzsis et al., 2001). Favorable temperatures for specific reactions could have been encountered in environmental niches, such as in local ice caps or hydrothermal systems.

3.3.4.3. Sources of catalysis. In all modern life-forms, enzymes lower the activation energy of numerous biochemical reactions necessary to sustain the living cell. In a world of prebiotic chemistry, none of those enzymes would have been available; hence important steps in the origin of life would have had to take advantage of a nonbiological catalytic mechanism or arise with the slower kinetics of uncatalyzed reactions.

Nonbiological catalysts that would have been available on prebiotic Earth include naturally occurring minerals such as clays (Meng et al., 2004; Ferris, 2005), zeolite (Smith, 1998, 2005; Munsch et al., 2001), sulfides (Wächtershäuser, 1988, 1990; Cody, 2004; Roldan et al., 2015), and calcite (Hazen et al., 2001; Hazen and Sholl, 2003). Other mechanisms that facilitated reactions could have been the concentration of reactants in compartments or on surfaces, and evaporation. Examples of concentrating processes include progressive shrinkage of brine pockets in sea ice during freezing (Kanavarioti et al., 2001; Price, 2007, 2009; Pierre-Alain and Hans, 2008), dehydration of pore spaces in subaerial beach sand or vesicular volcanic rocks, enrichment of hydrophobic organics at the ocean-atmosphere interface (Sieburth, 1983), or encapsulation in porous carbonate or pyrite precipitates that form near alkaline and acidic hydrothermal systems, respectively (Russell and Hall, 1997; Baaske et al., 2007). All these mineralogical and hydrological subsettings would have exchanged reactants and products via various mixing processes, which would include ocean and atmospheric circulation; thermal convection in the subseaﬂoor; tidal interaction; and catastrophic events such as impacts, seismic activity, or volcanic eruptions.

3.3.4.4. Causes of stereoselectivity. Stereoselectivity is a key feature of many modern biological reactions (see Chapter 4), but it is unknown at what stage in the origin of life stereoselectivity arose. There are numerous studies on obtaining stereoselective (enantiomerically enriched) products in putative prebiotic reactions between monomers (Popa, 1997). Even the smallest possible enantiomeric enrichment could be significant in prebiotic synthesis, as it could build up over time or in subsequent reactions (Mason, 1991). Some, however, consider that, on prebiotic Earth, obtaining any enantiomerically enriched products was impossible, and the abiogenic synthesis was either achiral or occurred elsewhere (Salam, 1991).

In mixtures of products from processes that researchers agree are abiotic (i.e., do not involve life), such as syntheses on comets and in interstellar clouds, some enantiomerically enriched compounds can be found (Pizzarello, 2006). Some inorganic processes, such as polarized light (Balavoine et al., 1974) or mineral catalysis (Kavasmaneck and Banner, 1977), are claimed to be responsible for enantiomeric enrichment of prebiotic syntheses products. When some chiral asymmetry is introduced to the pool of substrates or catalysts, it can be amplified in the subsequent reactions, resulting in almost enantiomerically pure mixtures (Soai et al., 1995; Blackmond et al., 2001).

3.3.4.5. Yields. Reactions between plausible prebiotic monomers in mixed media generate a huge variety of products; the yields for specific, biochemically “useful” compounds are therefore usually much lower than in controlled experiments with fewer reagents. However, given the fact that prebiotic processes likely had a huge reaction volume and many hundreds of years to accumulate products, single-digit percentage yields may have been sufficient. Also, in prebiotic processes, several analog compounds often have similar
activity. Often prebiotic reactions result in a whole class of compounds, different species with a certain functional group, rather than a single well-defined product structure.

The low yields and lack of selectivity are among the major problems of prebiotic chemistry (Sutherland and Whitfield, 1997). Modern biochemical processes, and most probably those of the first cell metabolism, are very selective. This high selectivity is achieved by using advanced enzymatic machinery. One of the biggest challenges of prebiotic chemistry (synthesis of building blocks) and protobiology (assembling the simplest biochemical reaction cycles) is to control the selectivity of the syntheses. One possible solution may be cyclical exchange of reagents and products between environmental niches that each favored and enhanced the abundance of very specific compounds (Stüeken et al., 2013). A recently highlighted example of such a niche is the pore space of rocks that preferably adsorb long over short oligonucleotides (Kreysing et al., 2015).

3.3.5. What processes could have formed the first organic carbon molecules? Prebiotic chemistry is the study of chemical reactions that result in the formation of organic carbon molecules, without the involvement of biological catalysts. This covers chemical reactions that take place under conditions such as those of the primordial Earth environment or the interstellar medium, where organic molecules can be synthesized from inorganic building blocks (Guillemin et al., 2004). Organic carbon molecules, such as precursors to lipids, nucleotides, amino acids, and carbohydrates, were necessary building blocks for the origin of life, but there is little agreement as to what the principal source would have been (McCollom, 2013; Trainer, 2013).

One of the first origin-of-life experiments, the Miller-Urey experiment (Miller, 1953), produced amino acids and other organic compounds by passing an electric charge through a mixture of gases that were considered to be components of Earth’s early atmosphere at the time of the experiment (H₂O, CH₄, and H₂, see Section 3.2). Impactors and interplanetary dust could also have delivered organic material (Chyba and Sagan, 1992; Pizzarello and Shock, 2010; Schmitt-Kopplin et al., 2010) and phosphorus (Pasek and Lauretta, 2005). The chemical processes that occur in hydrothermal vents, particularly those driven by serpentinization, are known to produce organic compounds through abiotic reactions. This could have resulted in the abiotic production of formate, acetate, methane, and larger hydrocarbons, including potentially amino acids, on early Earth (McCollom and Seewald, 2007; Proskurowski et al., 2008; Lang et al., 2010; Huber et al., 2012). Vents of this type may have been common on early Earth if oceanic crust was richer in olivine (e.g., Nna-Mvondo and Martinez-Frias, 2007) and if volcanic and tectonic activity was more widespread than today (e.g., Hargraves, 1986; Martin et al., 2007). Many simple organic compounds can be obtained with high yields in redox reactions catalyzed by iron, which was present in prebiotic, non-oxidizing oceans in much higher concentrations than in present-day oceans (Derry and Jacobsen, 1990) and might have served as a redox catalyst on prebiotic Earth (Caïns-Smith, 1978).

Ammonia, as a source of nitrogen for prebiotic synthesis of amino acids and amines, is readily synthesized by iron-catalyzed reduction of inorganic nitrates in simulated prebiotic ocean conditions (Summers and Chang, 1993). Amino acids and other simple carboxylic acids can also be synthesized in iron-catalyzed redox reactions (Schoonen et al., 2004). Various methods of prebiotic amino acid synthesis have been described (Pascal et al., 2005), and amino acids were products of the first successful prebiotic chemical synthesis experiment (Miller, 1953).

One of the oldest known methods of amino acid synthesis, the Strecker synthesis (Fig. 8), can yield enantiomerically enriched products when catalyzed by chiral catalysts (Li et al., 2003). Although currently known catalysts are not likely to have been available under prebiotic conditions, it is not impossible that some other types of chiral catalysts would have been used in the Strecker syntheses on primordial Earth.

It has been proposed that racemic mixtures of amino acid alkyl and thio-alkyl esters that self-assemble on the water/air interface can undergo spontaneous condensation, forming randomly sequenced peptide chains (Zepik et al., 2002). However, recent studies show that, in the case of short side chain natural amino acids, the reaction does not proceed beyond the stage of di- and tripeptides (Eliash et al., 2004), so it cannot be regarded as a prebiotic way of obtaining enantiomerically pure peptides and selecting amino acid products of a specific chirality.

A readily available source of carbohydrate derivatives on prebiotic Earth was reactions in the atmosphere that resulted mainly in formaldehyde formation. This simplest aldehyde can be formed under a variety of conditions, including UV irradiation of the primordial atmosphere composed of H₂O, CO₂, and H₂ (Pinto et al., 1980) or even simply H₂O and CO₂ (Bar-Nun and Hartman, 1978); electric discharge, possible in a wide range of atmospheric compositions based on CH₄, H₂, H₂O, N₂, and CO or CO₂ (Miller and Schlesinger, 1984); or photoreduction reactions (Chittenden and Schwartz, 1981). Various other aldehydes and sugars can be synthesized by condensation of formaldehyde (Weber, 1992) or by direct synthesis pathways similar to those of formaldehyde formation. Formaldehyde is also a substrate in the formose-type reactions—the sugar polymerization—which has been proposed as a historically first prebiotic route to obtain various mixtures of sugars (Breslow, 1959; Fig. 9).

It has been shown that amines (as amino acid derivatives or free primary amines) also can effectively catalyze sugar backbone formation (Weber, 2001). Long-chain carboxylic acids can be synthesized in Fisher-Tropsch-type syntheses, producing aliphatic chains and carbonyl groups in the inorganic gas condensations catalyzed on the mineral beds.

**FIG. 8.** Amino acid synthesis in the Strecker reaction.
Several methods of prebiotic synthesis of heterocyclic nucleobases, pyrimidines, and purines are known (Pavey et al., 1995; Shapiro, 1999). All four nucleobases can be synthesized from formamide, in the reaction catalyzed by various inorganic mineral catalysts (Saladino et al., 2001). Cyanoacetaldehyde condensation, catalyzed by guanidine hydrochloride, is a possible pathway to pyrimidines, and starting from thiourea, a thiopyrimidine product can be obtained (Robertson et al., 1996). Hydrogen cyanide condensation in the presence of ammonia yields purines and purine derivatives (Lowe et al., 1963).

Phosphorus was probably introduced into prebiotic synthesis at an early stage of prebiotic evolution (Macia et al., 1997). Phosphorus can be obtained from various inorganic sources (Brown and Kornberg, 2004). For example, phosphonic acid, synthesized from sodium phosphite (Fig. 10), was found in the Murchison meteorite (De Graaf et al., 1997). In fact, extraterrestrial delivery of the phosphide mineral schreibersite may have introduced a highly reactive form of phosphorus into prebiotic chemistry (Pasek et al., 2013, 2007).

A separate branch of prebiotic chemistry research is focused on the problem of polymerization of previously synthesized building blocks. Various studies have shown that amino acids can randomly polymerize to peptides (Plankensteiner et al., 2005), and nucleobases can form nucleic acid chains (Sutherland and Whitfield, 1997) under prebiotic, template-free conditions. The exact mechanism and conditions of such polymerizations are now the subject of extensive study.

All major organic functional groups can therefore be synthesized prebiotically; there are known reactions that can lead to the formation of amines, carboxylic acids, aldehydes and ketones, and heterocycles. All key components of the metabolic reactions can be synthesized, in various pathways leading to the amino acids, carbohydrates, lipids, and nucleobases. Also, inorganic sources of phosphorus and sulfur can be used to introduce these atoms to the organic compounds (Fig. 11).

3.3.6. What are the proposed sequences of events in the origin of life? Prebiotic Earth provided many different sites for possible prebiotic chemical reactions, including the open water of the prebiotic ocean, lagoons, surfaces of various minerals, thin layers of organic compounds, the gaseous phase of the atmosphere, or submarine hydrothermal vents. Different prebiotic processes proposed in the literature are placed in different conditions.

The origin of life on Earth might have not been a singular event. There is no reason to assume that our cells’
metabolism represents the only possible type of metabolic process; there could have been multiple origins of life. The evidence suggests that all known life comes from a single ancestor (see below), which implies that if there were multiple origins of life, only one lineage succeeded and survived, proliferating into all known forms of life. This LUCA—the last universal common ancestor of all organisms (Penny and Poole, 1999; Delaye et al., 2005)—was most likely a population of cells resulting from previous prebiotic, and earliest biological, evolution and selection.

It is not impossible that origin-of-life processes are still occurring, although they are far less likely to occur in the modern oxidized environment. Further, on a planet already inhabited it would be extremely difficult for any other form of metabolism to grow well enough to compete with “our type” of life. No effective biogenesis processes have been observed to date (Delaye and Lazcano, 2005).

Research into the sequence of events that led to the origin of life has focused on three essential aspects of modern life: metabolic networks, self-replicating genetic systems, and compartmentalization of the metabolites. Individual researchers tend to focus on the origin of one of these three features as the most essential event in the origin of life, but these debates are influenced by differences in each author’s implied definition of life (see Chapter 2). In particular, origin-of-life research is often characterized as a debate between the RNA-first model and the metabolism-first model; the RNA-first model proposes the origin of self-replicating genetic systems before the metabolic networks, and the metabolism-first model proposes the origin of metabolism before the rise of genetic systems.

Debates arise when authors ascribe the status of “life” to one of these systems but not the other. Nevertheless, it is clear that both metabolic networks and self-replicating genetic systems must have been integrated and encapsulated within a compartment before the evolution of the first modern living cell could have started.

3.3.6.1. Metabolic networks. One of the most fundamental observations in biology is that all forms of life share the same core metabolic pathways and biochemical building blocks. This profound fact of life, known as the “unity of biochemistry,” implies that the core metabolism must be ancient and date back to LUCA (see Chapter 4).

Historically, research into the origins of metabolism has been structured in terms of the “organic soup” model that was inspired by the famous Miller-Urey experiment (Miller, 1953) as well as Darwin’s vision of a “warm little pond.” Investigators have employed theoretical models, some quite mathematically sophisticated (Kauffman, 1993), to attempt to explain how a shallow pool full of organic compounds could react with each other in such a way that their interactions as well as their chemical compositions increase in complexity until a primitive metabolism is achieved. Little experimental evidence exists, however, for an organic soup that leads to reactions resembling a metabolic network.

The “organic soup” model implies present-day metabolism was a subset of a nearly infinite array of reactions...
among a random collection of many different kinds of starting materials. Alternatively, it is possible that the steps that led to proto-metabolic networks were more constrained (Morowitz and Smith, 2007). In this view, present-day metabolism is a direct consequence of prebiotic chemistry, rather than an accidental by-product, and the origin of RNA and DNA is intimately linked with these proto-metabolic reactions (Copley et al., 2007).

Support for the constrained network model is also derived from recent studies of the links between metals and the catalytic properties of enzymes in modern metabolic pathways. Carbon fixation pathways, in particular, appear to be ancient and involve a remarkable number of enzymes that require metal cofactors in their catalytic centers. Experimental studies have shown that the metals alone, without any organic components, can catalyze many of the reactions that constitute modern carbon fixation pathways (Huber and Wächtershäuser, 1997; Cody et al., 2000, 2004). Furthermore, the metals involved in these reactions are typically found in abundance at hydrothermal vent systems. The similarities between naturally occurring minerals and carbon fixation enzymes are unlikely to be coincidences, and an emerging model is that the first biochemical pathways evolved based on already operating geo-chemical reactions (Cody and Scott, 2007; Nitschke et al., 2013; Russell et al., 2013, Sousa et al., 2013).

Peptide-based metabolic networks have also been studied (Goldman et al., 2012). Recent evidence suggests that small, prebiotically plausible peptides could have been catalyzing nucleic acid synthesis and peptide synthesis before the origin of protein catalysts (Li et al., 2000; Gorlero et al., 2009; Wieczorek et al., 2012; Adamala and Szostak, 2013a). For example, the simple dipeptide catalyst seryl-histidine (Ser-His) can drive vesicle growth through the catalytic synthesis of a hydrophobic dipeptide, N-acetyl-l-phenylalanine leucinamide (AcPheLeuNH₂), which localizes to the membrane of model protocells and drives competitive vesicle growth in a manner similar to that previously demonstrated for phospholipids. Ser-His has previously been shown to catalyze the formation of peptide bonds between amino acids and between peptide nucleic acid (PNA) monomers (Gorlero et al., 2009). Although Ser-His is a very inefficient and nonspecific catalyst, Ser-His generates higher yields of peptide product in the presence of fatty acid vesicles. As a result, vesicles containing the catalyst generate sufficient reaction product to exhibit enhanced fitness, as measured by competitive growth, relative to those lacking the catalyst. It is therefore possible to observe how a simple catalyst causes changes in the composition of the membrane of protocell vesicles and enables the origin of selection and competition between protocells (Adamala and Szostak, 2013a).

3.3.6.2. Genetic systems. The traditional central dogma of molecular biology states that genetic information always flows one way from DNA through RNA to protein. DNA acts as information storage. RNA acts as messenger. Protein carries out catalytic functions. The elucidation of the central dogma immediately created a difficult “chicken and egg” problem for the origin of life. If DNA originated first, how did it replicate itself without proteins? But if proteins originated first, how was their informational content encoded?

The RNA World model (Gilbert, 1986) for the origin of life is extremely popular because of its elegant solution to this quandary: RNA acted as both a genetic information-storing molecule and a catalytic molecule. Today, RNA serves as the sole genetic molecule in many viruses, and the number of reactions catalyzed by RNA enzymes (“ribozymes”) is ever increasing (Doudna and Cech, 2002; Chen et al., 2007). Therefore, one can envision a primitive cell in which genetic information is encoded by RNA and functions are carried out by ribozymes.

According to the RNA World hypothesis of the origin of life on Earth, all reactions in protocells were catalyzed by nucleic acid enzymes before the origin of protein-based catalysis known from modern biology (Gilbert, 1986; Szostak, 2012). The RNA enzymes can catalyze a wide variety of reactions, encompassing all types of transformations necessary for the protocell activity (Doherty and Doudna, 2001; Lilley, 2003). It is unclear, however, how the very first ribozymes came into existence in the absence of catalysts capable of synthesizing long, functional RNA.

Non-enzymatic RNA polymerization allows synthesis of short RNA oligonucleotides under prebiotic conditions (Joyce et al., 1984; Adamala and Szostak, 2013b). In this reaction, the growing RNA primer strand is elongated by addition of RNA nucleotides activated with a good leaving group, for example, an imidazole or imidazole derivatives (Hey and Göel, 2005; Szostak, 2012). This reaction, while giving a prebiotically plausible route to RNA oligos on early Earth, is not efficient enough to allow for high-yield synthesis of long, active RNA enzymes (Fig. 12).

The RNA World hypothesis states that RNA must have served important genetic and catalytic roles in life’s early stages. On the other hand, some authors stress that a complex and unstable molecule like RNA could not have existed on its own without a supporting metabolic network composed of simpler molecules. Indeed, the intricate and universal relationship between RNA and amino acids that led to the genetic code seems likely to have been a very early development in the evolution of life (e.g., Copley et al., 2005).

While the RNA World hypothesis remains very popular with many astrobiologists, some experiments suggest that earliest synthesis and catalytic activity of RNA could have been supported, or supplemented, by different molecules. Other models for the origin of the first genetic system include systems based on peptides (Rode, 1999) or clay minerals (Cairns-Smith and Hartman, 1986).

The prebiotic replication of RNA could have occurred non-enzymatically inside protocells via chemically activated nucleotides (Joyce et al., 1984; Mansy et al., 2008; Adamala and Szostak, 2013b). The coupling of lipid membrane compartmentation (see Section 3.3.6.3) with the possibility of RNA activity and replication, and the division of the protocell compartment, provides a scenario for the possible origin of the interaction between the RNA World and protocell compartments, leading to the formation of the first evolving protocell.

3.3.6.3. Encapsulation. All living cells are encapsulated by lipid membranes that serve essential biological functions including transport, communication, and energy conservation. For example, all organisms today conserve energy for metabolism by establishing a chemiosmotic gradient across their lipid membranes. The prebiotic origin of the lipid membrane is still not fully explained. The open question about
the nature of the membrane in LUCA leaves many possible routes to the origin of lipid membranes during the earliest stages of protobiological evolution.

In modern cells, apart from compartmentation, membranes perform several other functions, including energy transduction and transport of organic and inorganic compounds, and provide a docking site for many enzymes. Presumably, the very first role of the membranes was simple encapsulation— isolation of the reaction cycles (i.e., genetic materials or enzymatic peptides) from the environment. This could be done by the simplest amphiphiles, possibly available under the prebiotic conditions: medium-sized (up to C10) chain carboxylic acids.

The main building blocks of modern cell membranes are phospholipids and sterols. Phospholipid glycerol esters and sterols are too complex to be synthesized under abiotic conditions. However, all these compounds can be derived from simple building blocks—sterols from isoprene units, and lipid derivatives from simple unsaturated carboxylic acids. Compounds based on these simplest units could have formed the first membranes that encapsulated biochemical cycles of the protocell (Walde, 1999).

In water, the simplest amphiphiles spontaneously self-organize into bipolar membrane sheets that close into spherical structures known as vesicles (Apel et al., 2002). Lipids in solution form vesicles spontaneously because of their amphiphilic nature; that is, they have one hydrophobic end and one hydrophilic end (Pohorille and Deamer, 2009). Vesicles are commonly accepted as an approximation of the compartments of the earliest protocells (Walde, 2006). Amphiphilic molecules collected from the Murchison meteorite spontaneously form vesicles when placed in aqueous solution (Deamer and Pashley, 1989). The ease with which vesicles form causes many researchers to assume encapsulation occurred almost immediately after lipids became available, either via extraterrestrial delivery or abiotic synthesis (Pohorille and Deamer, 2009). Encapsulation may have strongly affected the dynamics of prebiotic chemistry by segregating molecules and reactions, increasing local concentrations, promoting coevolution of biochemicals, allowing selective permeability, and creating membrane gradients. It is not clear how large biochemicals would have become encapsulated initially, but several models have been proposed (Deamer et al., 2002).

As with the origins of metabolic networks, many researchers view encapsulation as mimicry of geochemical reactions that were already occurring on early Earth. Russell and Martin, for example, developed a highly detailed model in which the origin of life occurs within the iron-sulfide bubbles of hydrothermal vent chimneys (Russell and Martin, 2004). Clays are also known to accelerate the formation of vesicles (Szostak, 2001). Vesicle structures can grow (Chen and Szostak, 2004), divide (Hanczyc et al., 2003), and selectively take up compounds from the environment (Chen et al., 2004). Therefore, investigating properties of the different vesicle systems can give insight into possible routes to the origin of protobiological compartmentalization. The origin of a selective membrane transport system was crucial for the origin of the living cell’s homeostasis. Self-sustaining dynamic equilibrium between the cell’s interior and the external environment required simultaneous development of two elements: the cell membrane and channels to control the permeability of the membrane. Most known natural ionophores are peptides; therefore, the prebiotic synthesis of peptide-based ion channels and pores may have been an important stage in the earliest, chemical evolution of life.
Chapter 4. How Have Earth and Its Biosphere Evolved?

4.1. What does the tree of life tell us about how life has evolved?

4.1.1. What is a phylogenetic tree, and how is it constructed? The “tree of life,” a diagram that organizes all living biological organisms, is designed in a way that allows us to understand the evolutionary relationship and shared features of those organisms (Gaucher et al., 2010). It demonstrates that all known life has evolved from a single origin, which could have been a single organism or a group of interacting organisms (Woese, 1998).

The tree of life is an example of a phylogenetic tree, which is a diagrammatic representation of how closely related genetic sequences are to each other. Any gene can be used to make a phylogenetic tree, and trees can be made by using DNA, RNA, or amino acid sequences. Different methods for calculating relatedness may be used and can give different resulting trees (see below). Different organisms have different types of genes, so the choice of which gene to use determines which organisms can be included in the tree; any organism that does not possess the relevant gene cannot be included in the tree. For the tree of life, which is intended to cover all life on Earth, the small subunit ribosomal RNA (ssu rRNA) gene is used to make the tree, as this is a gene that is essential to cellular processes, so there is very good reason to believe that all cells possess it (see Section 4.1.2). The degree of relatedness of the ssu rRNA gene is used to define taxonomic groups, which are groups of organisms that are more closely related to each other than to organisms outside the group at different levels. For example, for microbes, 97% identical is generally considered equivalent to a single species, while 95% identical is considered equivalent to a single genus, although no exact percentage is currently agreed upon to define lower-level taxonomic positions. The ssu rRNA tree is therefore commonly considered to demonstrate relationships not just between a single gene but between different organisms as well. One important caveat to this is that, due to horizontal gene transfer (see Sections 2.3 and 4.1.3), it is possible that all the genes in an organism may not have the same evolutionary history. Viruses are not cells, do not contain the ssu rRNA gene or an equivalent, universally conserved genetic feature, and so are not included in the ssu rRNA tree, and there is no consensus as to whether they should be considered alive (see Sections 2.3 and 4.1.2).

To make a phylogenetic tree, genetic sequences are compared; organisms with closely related sequences are located in close proximity, and organisms with distantly related sequences are located far apart on the tree (Fig. 13). A phylogenetic tree therefore provides information about the progress of evolution by showing the relative order in which the DNA sequence of a gene has changed over time. Prior to the advent of genetics, the relatedness between two different organisms was determined by comparing physiological characteristics and physical appearance. However, organisms with similar characteristics (such as similarly shaped beaks for birds) may not have a common ancestor if they obtained those features from convergent evolution instead of divergent evolution. In divergent evolution, the organisms inherited the characteristic from a common ancestor, and it changed into slightly different forms in each descendant branch. In convergent evolution, two different groups of organisms evolved the same characteristic completely independently. Phylogenetic trees can tell us whether organisms with shared physiological characteristics have a common ancestor or whether those traits evolved independently. However, it is important to understand that the vast majority of sequences on a tree come from organisms that are currently alive (although extinct organisms can be included if DNA is available from them); these are the tips of the branches of the tree. Historical information about how the gene sequence has changed is deduced from the structure of the tree—including where the nodes, or branch points, are—which is dependent on the method used to construct it (Fig. 13). Different methods include the rectangular phylogram, the rooted rectangular cladogram, and the bifurcating unrooted tree. The first two of these have a root, which specifies the most evolutionarily ancient branch. The difference between these two is that branch lengths on the phylogram are proportional to evolutionary distance, whereas all branches are equal in length on the cladogram. The third tree shown here—the bifurcating unrooted tree—is unrooted and used when it is not clear which branch is the most ancient.

There are various methods used in the construction of a tree. Neighbor-joining (a distance method) is probably the simplest method. In this method, a distance matrix is calculated that shows how closely related the sequences are to each other, based on how much of the sequence is identical (the percentage identity) and how much is different. The distance matrix is then used to construct the tree diagram as a simple map of the distances between each sequence (just as in the same way a matrix of distances between different cities in a country could be used to map their geographic relationship to each other). The neighbor-joining method is fast to calculate because it only uses one value to assess the difference between each pair of sequences. Character-based methods take each character into account instead of just using one overall value. Bayesian, maximum parsimony, and
maximum likelihood are among the most commonly used character-based algorithms. A detailed comparison of maximum likelihood and Bayesian approaches can be found in the work of Archibald et al. (2003).

Recently, some astrobiologists have given consideration to the length of the branches on the tree-of-life phylogeny and have attempted to scale phylogenetic trees with absolute time. These efforts are often referred to as establishing “molecular clocks” that can keep track of the timing of branch points on phylogenetic trees. The goal of such work is to further our understanding of evolutionary events in the fossil record through an analysis of genetic or amino acid substitutions during the evolutionary history of a lineage. To this end, Hedges and Kumar (2009) argued for aligning biological time measurements, such as the fossil record, to phylogenetic trees to create “timetrees” that establish absolute time of divergence between species.

However, the dates of divergence between taxa—the time that has elapsed since two species shared a common ancestor—are more difficult to establish than the phylogeny (Roger and Hug, 2006). It is only possible to assign an absolute date to a phylogenetic change if there is additional evidence of the date from the fossil record. It is also clear that the rate of evolutionary change is not constant. There is variation in the genetic sequence substitution rate among generations and between evolutionary lineages, which means that different branches of the tree of life have evolved at different rates (Pulquerio and Nichols, 2007). The effects this variation could have on molecular clocks, together with frequent disagreement between the ages estimated by way of the fossil record and molecular clocks—especially for the most ancient lineages (Hedges et al., 2004; Donoghue and Benton, 2007)—mean that it may not be possible to produce molecular clocks that can constrain the uncertainty enough to provide any useful information. For these reasons, many researchers do not consider molecular clocks to be a reliable technique (Heads, 2005).

4.1.2. What are the three domains of life? Life on Earth is divided into three domains: Archaea, Bacteria, and Eukarya (or Eucarya). Single-celled organisms are common in all three domains, and all organisms in the domains Archaea and Bacteria are single-celled. All multicellular organisms fall within the domain Eukarya. The term “prokaryotes” has traditionally been used to describe single-celled organisms that do not contain a nucleus, organelles, or a cytoskeleton (collectively, the domains Bacteria and Archaea). However, further work has led some biologists to completely abandon the term “prokaryote,” as it implies that the Bacteria and Archaea are more closely related to each other than to the Eukarya, which is now known to be false (Pace, 2006). See Fig. 3 in the work of Pace (2009, at http://mmbr.asm.org/content/73/4/565/F3.expansion.html) for a diagram of the tree of life.

The current three-domain classification was proposed by Carl Woese and colleagues in 1977 (Woese and Fox, 1977) and later refined in 1990 (Woese et al., 1990), based on the nucleotide sequence of the gene for the small ribosomal RNA subunit (“16S” rRNA for Bacteria and Archaea and “18S” rRNA for eukaryotes). All living organisms contain ribosomal RNA, which plays an essential role in protein synthesis (see Chapter 3). Given its ubiquitous presence, researchers frequently utilize ssu rRNA gene sequences to determine the evolutionary relationships among organisms. Such utilization for determining phylogeny depends on an assumption of very low impact from horizontal gene transfer (the transfer of genes between two distinct species rather than by descent from a parent organism; Choi and Kim, 2007; also see Sections 2.3 and 4.1.3). The slow evolution of the ssu rRNA gene makes it an excellent choice when comparing diverse species but less useful at differentiating close relatives (Pace, 2009). Additional genes, including entire genomes, may be used when this becomes an issue.

The majority of life on Earth is microbial, and astrobiologists extrapolate from that fact to conclude that the...
majority of any extraterrestrial life that may possibly exist is also likely to be microbial. Archaea and Bacteria are the focus of astrobiological studies due to the diversity of metabolic strategies utilized (see Section 4.2.1), their ability to thrive in extreme habitats (Pikuta et al., 2007), their pivotal role in Earth’s biogeochemical cycling, and the fact that single-celled organisms dominated the early history of life on Earth. Bacteria and Archaea harbor the vast majority of the genetic diversity of life on Earth. Scientists estimate that, to date, less than 1% of microbial life has been success-

fully cultivated (e.g., grown in the laboratory) (Rappé and Giovannoni, 2003). Therefore, molecular methods that do not require cultivation are also used to detect the presence and genetic capabilities of microbes through DNA (metagenomics), RNA and gene expression activity (meta-

transcriptomics), and transcribed proteins (metaproteomics). These techniques have demonstrated that microbial populations are highly complex and that we have only begun to elucidate the full extent and capabilities of microbial life (DeLong, 2009).

An active community of microbiologists also considers viruses to be a form of life (Forterre, 2011). Viruses do not bear the ssu rRNA genes used by Woese and colleagues in the construction of the current tree and designation of the three domains. If viruses are classified as a life-form, the tree of life would need to be fundamentally restructured, as viruses are incredibly abundant and diverse, numerically outnumbering cells by a factor of 10 (Dinsdale et al., 2008). Viruses also serve as mobile genetic elements, moving functional genes between different species and even between different do-

mains, through a process called transduction, a process driven by their coevolution with the host organisms. This pheno-

menon of genetic acquisition is anticipated to have played, and is expected to continue to play, a paramount role in both the origin and evolution of life.

4.1.4. What do we know about the possible nature of a last universal common ancestor? LUCA is the last universal common ancestor from which all modern life descends. It is not the most primitive cell or the first cell that ever existed; however, it is the universal common ancestor of all existing organisms. It may not have been a single organism but a group of cells or protocells that evolved together (Woese, 1998). The characteristics of LUCA are difficult to identify due to the extremely long period of time that has passed since its existence. Although the details require further characterization, researchers have begun to probe the properties of LUCA by identifying shared characteristics of deeply rooted organisms. LUCA is estimated to have lived 3.8–3.5 billion years ago (Glansdorff et al., 2008). Phylogenetic analyses, including whole genome sequencing, reconstruction methods, and homology analyses, suggest that LUCA had genes that are required for a minimal organismal function (Mat et al., 2008). Other analysts suggest that LUCA was a single-celled organism with a lipid membrane and a DNA-based genetic code with 500–1000 protein-coding genes, capable of RNA biosynthesis, and had its own translation machinery to assemble proteins from

generation (e.g., ATPase) and production of proteins (e.g., the ssu rRNA gene) produce a tree-of-life hierarchy with the same resolution of three domains, suggesting a common core of genes that have evolved together, even though other genes have different evolutionary histories. Similar domain-resolving phylogeny indicates that the root of the tree of life lies within the bacterial domain and that the Archaea and Eukarya share a common ancestry after their common lineage separated from the bacterial line of descent (Ciccarelli et al., 2006).

One explanation for the diversity of evolutionary histories of genes within a single organism is the phenomenon known as horizontal (or lateral) gene transfer. While genetic material is classically thought of as vertically derived (from one’s progenitors), enormous amounts of genetic diversity are conferred laterally, from sources other than one’s progenitors (McDaniel et al., 2010). Horizontal gene transfer can take place by different mechanisms, as follows: some bacteria will take up DNA (that, for example, could have been released by the death of another cell) directly from the environment, DNA can be transferred directly from one cell to another, or DNA can be transferred by viruses as they infect different types of cells.

While the extent of the impact of horizontal gene transfer upon the classification of organisms and phylogeny is still poorly constrained, horizontal gene transfer among the three domains has led some researchers to reframe the tree of life as a “reticulated tree or net” (Doolittle, 1999), a “cobweb of life” (Ge et al., 2005), and a “coral of life” (Fournier et al., 2009; Gaucher et al., 2010), each name reflecting the fact that the evolutionary history of life is not strictly linear.

While phylogenetic trees produced on the basis of genomic data can provide evolutionary insight, they have numerous limitations. Debate continues about the nature of the tree of life and its root (e.g., Lake et al., 2008). The tree-of-life concept itself is a topic of active discussion in light of molecular genetics and the microbial species concept (O’Malley and Koonin, 2011), and as a result, resolving the nature of LUCA remains obscured.
4.2. How has life coevolved with the surface of the planet?

4.2.1. How does life on Earth obtain energy from the environment? At the most fundamental level, life exists on chemical energy derived from the differing ability to attract and hold electrons in chemicals made up of H, C, N, O, S, Fe, Mn, and other elements (Levin, 2001; Falkowski, 2006; Falkowski et al., 2008). The processes by which electronic energy is exchanged are termed redox processes. In these processes, reduction entails becoming more electron rich (i.e., gaining electrons and becoming more negative), whereas oxidation refers to becoming more electron deficient (i.e., losing electrons and becoming more positive). In a similar way to the operation of electricity where positive and negative terminals are required for a functioning circuit, electrons in chemical reactions “flow” such that as one chemical species becomes reduced another becomes oxidized, and the electrons are transferred via molecules within the cell in a way that allows the cell to obtain energy from the redox reaction.

Organisms that use light as an energy source are called phototrophs. Light energy (photons) is also utilized by life through photosynthesis, and the underlying mechanisms that are used to harness this energy are also redox reactions. The dominant source of energy to Earth’s primary producers (organisms that fix carbon dioxide into organic carbon, so forming the bottom of the food chain) is thought to be oxygenc photosynthesis (Sleep and Bird, 2007). In oxygenc photosynthesis, photons are used to split H₂O into O₂, protons (H⁹⁺), and electrons. The protons and electrons are then used to create energy storage molecules, which in turn are used to reduce CO₂ into organic carbon. Thus, the “net” redox pairs involved in oxygenc photosynthesis are H₂O being oxidized to O₂ and CO₂ being reduced to organic carbon. There are also anoxygenic photosynthetic mechanisms that utilize photons for energy but do so with electron sources other than H₂O and create by-products other than O₂. For example, some organisms are capable of utilizing hydrogen sulfide (H₂S) as an electron source for photosynthesis, oxidizing the sulfide and producing S⁰ as a by-product. These variants are widespread, but these reactions do not yield as much energy as oxygenc photosynthesis.

Numerous organisms from both the bacterial and archaeal domains are capable of exploiting the redox potential of chemical species produced by biotic and abiotic (nontubiological) sources. The potential energy present in the pairing of any two chemical species is best represented in an electron tower (Fig. 14), where more reduced species are found near the top of the tower. Any of the pairings represented on the tower can provide energy for growth, including the couple all humans use: organic carbon with oxygen as an electron acceptor. Those organisms that derive energy from chemical species instead of light are called lithotrophs if they use chemical energy sources other than organic carbon, or they can be called chemotrophs, a more general term that covers any chemical energy source. Organisms may use the energy captured to reduce CO₂—using it to create organic carbon molecules (autotroph)—or they may use organic carbon molecules from other organisms (heterotrophy). Organisms can therefore be described in terms of both how they obtain energy and how they obtain carbon; for example, a lithoautotroph obtains energy from inorganic chemicals (litho-) and fixes carbon from CO₂ (autotroph). On Earth, virtually every environment that contains minimal redox potential and additional requirements for water activity and temperature (see Chapter 5) appears to host life. Thus, detecting potential sources of energy for life in extraterrestrial environments is a high priority of the astrobiology community (Hoehler et al., 2007).

4.2.2. How do life and global geochemical cycles mutually affect each other? Carbon (C) forms the basis of matter
and energy cycling in all known forms of life and is abundant in all organic biomolecules. The long-term C cycle on Earth operates over millions of years and is controlled by silicate weathering, tectonic activity, and sedimentation of the calcium carbonate shells of marine organisms. The short-term C cycle is mediated by biology; the greatest flux of C occurs through photosynthetic fixation of CO$_2$ (see Section 4.2.1), which is estimated to convert over 10% of atmospheric CO$_2$ ($\sim 10^7$ g) into biomass annually (Schlesinger, 2013). Diverse chemoorganotrophs (including humans) reoxidize this organic C to CO$_2$ to gain energy. Beyond its role as a bioessential element, C is also important for regulating global climate due to the negative feedback loop of the carbonate-silicate cycle; increased CO$_2$ results in warmer temperatures, more weathering, enhanced limestone formation, and an increased removal of CO$_2$ from the atmosphere that offsets the original increase. Over millions of years, this feedback mechanism has helped stabilize global climate and is believed to have led to climatic recovery from past “Snowball Earth” events (Hoffman and Schrag, 2002).

Nitrogen is present in amino acids, nucleic acids, porphyrins, and other biomolecules. The great abundance of nitrogen (N$_2$) in Earth’s atmosphere is derived from volcanoclastic outgassing from Earth’s mantle. The only known biological mechanism by which N$_2$ gas is made bioavailable is through enzymatic reduction to ammonium (NH$_4^+$) by some nitrogen-fixing bacteria and archaea. The rise of atmospheric O$_2$ during the Great Oxidation Event at 2.4 Ga enabled the production of nitrate (NO$_3^-$) through the microbial process of nitrification (the oxidation of NH$_4^+$ to NO$_3^-$) (Falkowski and Godfrey, 2008). Higher levels of N$_2$ in the Archean atmosphere could have contributed to an enhanced greenhouse warming that helped offset the lower heating provided by the faint young Sun (Goldblatt et al., 2009). During the mid-Proterozoic, this greenhouse warming might have come from nitrous oxide (N$_2$O) gas (Buick, 2007).

Oxygen (O) is essential as a building block for organic molecules, and when present as molecular O$_2$, it is the most thermodynamically favorable electron acceptor for aerobic respiration. Oxygen cycling on Earth has also dramatically changed in response to biological evolution. While O in the form of liquid water has been continually present on Earth’s surface since $\sim$ 200 million years after the accretion of the planet (Falkowski andIsozaki, 2008), O$_2$ gas was essentially absent from the atmosphere until the Great Oxidation Event, when O$_2$ concentrations at the surface of the planet rose from less than a part per million to a few percent of the atmospheric O$_2$ (Farquhar, 2000; Canfield, 2005; Holland, 2006; Lyons et al., 2014). This appearance of O$_2$ in the atmosphere also led to the development of an ozone (O$_3$) layer and an O$_2$/O$_3$ UV shield that protected land-based life from damaging radiation. Ultraviolet shielding by O$_2$/O$_3$ becomes effective above $\sim$ 1% O$_2$ (Kasting and Donahue, 1980). The buildup of O$_2$ at the Great Oxidation Event was the result of cyanobacterial O$_2$ production after the advent of oxygenic photosynthesis. Organisms that could cope with some O$_2$ would have flourished after the Great Oxidation Event, since O$_2$ is a very effective electron acceptor for respiration when paired to a wide range of electron donors (see Section 4.2.1). A second rise in atmospheric O$_2$ to at least 3% of the atmosphere (at least 15% of the present atmospheric O$_2$ level) occurred in the Ediacaran ($\sim$ 560–550 million years ago, Ma). This O$_2$ rise was likely due to continental breakup and increased C burial. These rises in atmospheric O$_2$ concentrations may have paved the way for early animal diversification by providing a high-energy metabolism—the reaction of organic carbon with O$_2$—to maintain these organisms’ larger bodies. This could have spurred the Cambrian explosion of complex animal life ($\sim$ 545 Ma) by providing enough O$_2$ for animals to breathe (Berner et al., 2007). A third rise in O$_2$ to near present levels ($\sim$ 20% of the atmosphere) has recently been proposed to have occurred $\sim$ 400 Ma in the Devonian as a result of invasion of land by vascular plants (see Dahl et al., 2010, and references therein).

Phosphorous (P) has been a major limiter of marine and continental primary productivity over geological timescales. Phosphorus is required for the construction of nucleic acids, adenosine triphosphate (ATP, the molecule that cells use as energy currency), and phospholipids in cell membranes (Saltzman, 2005; Filippelli, 2008). In contrast to life’s other major elements, P does not have a major atmospheric gas phase. Instead, the major source of reactive P is weathering of calcium phosphate minerals (such as apatite) from rocks (Schlesinger, 2013). It is possible that small continental land masses in the Hadean and early Archean limited the flux of P to early oceans. Low P content in mid-Archean and early Proterozoic sedimentary deposits of iron oxides (banded iron formations) was originally taken as evidence of limited P availability during the Archean and early Proterozoic eras (Bjerrum and Canfield, 2002). However, the iron formation data have recently been linked to high dissolved silica concentrations rather than a P crisis; silica can outcompete P for binding sites on iron oxides (Konhauser et al., 2007; Planavsky et al., 2010). Iron formations linked to the widespread Neoproterozoic glaciations preserve a signal for an anomalously large marine P reservoir (Planavsky et al., 2010). Reduced P species have recently been suggested to react with prebiotic molecules on early Earth (Pasek et al., 2013). These high P concentrations could have also caused the Great Oxidation Event at $\sim$ 2.45 Ga by providing a large nutrient source for oxygenic photosynthesis, allowing for more biological productivity and the burial of more reduced organic carbon.

Sulfur (S) on the surface of Earth is the product of volcanic outgassing through time. In biology, S is utilized in amino acids (methionine and cysteine) in enzyme active sites. Because S can exist in several different oxidation states, sulfur compounds are metabolically abundant and are a major energy source for microbes (Beinert, 2000; Sekowska et al., 2000). With the Great Oxidation Event, increasing amounts of sulfate (SO$_4^{2-}$) would have been weathered into the ocean and utilized by anaerobic microbes as an electron acceptor to oxidize organic carbon (in anaerobic respiration) with the concomitant production of sulfide (H$_2$S) (Canfield, 1998). Sulfide could have sustained primary production through anoxygenic photosynthesis, with H$_2$S serving as the electron source (see Section 4.2.1; Johnston et al., 2009). Sulfur-containing metabolites can exhibit large global effects because dimethylsulfide produced primarily by marine organisms plays an important role in cloud nucleation and climate (Charlson and Warren, 1987), although this feedback loop is more complex than was originally recognized (Quinn and Bates, 2011).
In addition to the macronutrients (CHONPS) that form the basis of life’s biopolymers (i.e., DNA, proteins, lipids, etc.; see Section 5.1.1), life utilizes smaller quantities of many other elemental constituents. Trace elements that are essential to virtually all life include Na, Mg, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, Se, Mo, and Cl, whereas elements that are only used by some life-forms include B, Si, V, Ni, As, Br, Sr, I, Cd, Ba, Cr, and W (da Silva and Williams, 1991). Many of these trace elements are essential due to their presence as cofactors in enzymes. Each of these elements has its own biogeochemical cycles, although nearly all these cycles are understudied compared to the major element cycles. The concentrations of many trace elements, particularly metals such as Fe, Ni, Cu, Zn, and Mo, appear to have dramatically changed over time in response to the redox state of the oceans and the amount of H2S in the deep ocean (Zerkle et al., 2005; Anbar, 2008; Lyons et al., 2009).

In summary, geochemical cycles enable global energetic disequilibria (i.e., that there is energy to be harvested) and sustained biological material flux (i.e., metabolism) on Earth. Therefore, tectonics is often thought of as a planetary-level constraint to long-term habitability (see Chapter 5).

4.2.3. What is the effect of metabolism on a planetary atmosphere? The composition of a planet’s atmosphere is influenced by several processes, as follows: geophysical processes such as volcanism and tectonics, chemical processes such as photochemical and surface reactions, and astrophysical processes such as erosion by solar wind and cosmic rays. If life is present, biological processes can influence these cycles by adding new gases, depleting others, or otherwise altering their abundance. Living organisms can dramatically influence the geochemical evolution of their host planet, as has happened on Earth, and if the biological imprint on a planet’s atmosphere is significant enough, it can be detected remotely. Indeed, the remote study of the atmospheres of extrasolar planets may be the only means of detecting the presence of life on worlds beyond our solar system (see Section 7.2).

Modern Earth’s atmosphere consists almost entirely of gases participating in biological cycles. Nitrogen and oxygen make up over 99% of the modern atmosphere, and their production on Earth is dominated by biological activity. Most of the gases making up the remaining 1% are also biologically cycled, and many of these trace gases have significant impacts on our planet’s environment and climate, including CO2, CH4, H2, CO, OCS, H2S, (CH3)2S, N2O, NO, NO2, and volatile organics. Table 2 lists the metabolisms responsible for cycling some of the more abundant atmospheric gases.

While some of the gases abundant in the modern atmosphere are thought to have been in similar abundances on prebiotic Earth, some of them, most notably oxygen, are anomalous with regard to abiotic planetary geochemistry and are detectible evidence of the presence of life on our planet. The atmosphere of Earth varies greatly in composition from its terrestrial neighbors (Fig. 15) in large part because of biological activity.

The chemical disequilibrium observed in modern Earth’s atmosphere is also a hallmark of our planet’s biosphere. The Galileo spacecraft, when doing a flyby of Earth, demonstrated that this disequilibrium could be remotely detected via spectroscopy of the atmosphere of the planet; this is similar to plans to detect this disequilibrium on potentially habitable extrasolar planets (Sagan et al., 1993; see also Section 7.2.2). The results collected showed that our atmosphere is far out of steady-state chemical equilibrium in large part because of the constant biological cycling of gases. Atmospheric species such as oxygen and methane (and other reduced gases) should not coexist in appreciable quantities in an equilibrium atmosphere (Lovelock, 1965; Sagan et al., 1993), yet they do on our living planet.

4.2.4. The evolution of multicellular life. Unicellular (single-celled) microorganisms have dominated life on Earth for most of its history and continue to be major forces impacting the geochemistry of the planet. However, multicellular organisms, from fungi to humans, play a major role in the cycling of many chemicals and gases and have conspicuously modified the surface landscape of Earth. Multicellularity provides advantages including protection from predation and environmental hazards, enhanced mobility, and improved resource utilization.

There are two major modes by which multicellular organisms could have initially evolved: (1) ordinarily free-living cells aggregate in response to some external driver, or (2) cells remain stuck together following cell division (Grosberg and Strathmann, 2007).

Complex multicellularity involves functional differentiation and also occurs in many branches of eukaryotes. Agglomerates involving different unicellular species (e.g., bacterial biofilms) or even different domains are examples of cooperative cellular behavior (Dekas et al., 2009; Overmann

---

<table>
<thead>
<tr>
<th>Gas</th>
<th>% by volume in modern atmosphere</th>
<th>Biological pathways responsible for gas production</th>
<th>Biological pathways responsible for gas consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>N2</td>
<td>78.08</td>
<td>Denitrification, anammox</td>
<td>Nitrogen fixation</td>
</tr>
<tr>
<td>O2</td>
<td>20.95</td>
<td>Oxygenic photosynthesis</td>
<td>Aerobic respiration</td>
</tr>
<tr>
<td>CO2</td>
<td>0.034</td>
<td>Respiration, methanotrophy, fermentation</td>
<td>Photosynthesis, nonphotosynthetic autotrophy, acetogenesis, methanogenesis</td>
</tr>
<tr>
<td>CH4</td>
<td>0.0002</td>
<td>Methanogenesis</td>
<td>Methanotrophy</td>
</tr>
<tr>
<td>H2</td>
<td>0.00005</td>
<td>Fermentation, proton reduction, nitrogen fixation</td>
<td>Hydrogen-based lithotrophy, acetogenesis, methanogenesis</td>
</tr>
</tbody>
</table>

Modern atmospheric compositions from the work of Rakhecha and Singh (2009).
and Hallenbeck, 2010); however, the traditional definition of multicellularity requires that the involved cells be genetically identical. Complexity can also occur within cells; a major event in the evolution of eukaryotes was when organisms capable of engulfing other organisms acquired mitochondria (permitting aerobic respiration for energy) and chloroplasts (permitting oxygenic photosynthesis) (Koonin, 2010). Both of these metabolisms were evolved by Bacteria, and the ancestors of modern eukaryotes acquired them by engulfment, which resulted in an obligatory symbiosis (Margulis, 1971; Gould et al., 2008). Mitochondria and chloroplasts have genomes distinct from their host cells, with ssu rRNA genes (see Section 4.1.2) that group phylogenetically with Bacteria (α-Proteobacteria and Cyanobacteria, respectively) rather than their host cells.

Regardless of how multicellularity arose, the geological record of multicellularity (in the form of sponge biomarkers) extends as far back as the Marinoan glaciation around 635 million years ago (Ma) (Love et al., 2009), with records of small tubular microfossils extending back to 716 Ma (Bosak et al., 2012) and controversial reports extending as far back as 1.8 Ga (Nichols and Wörheide, 2005). Fossil sponges, as well as large, somewhat enigmatic fossils of complex eukaryotes with no known extant relations (known as the Ediacaran biota; Seilacher, 1984), are found in various locations around the world, dating from 585 Ma until the Cambrian explosion of animal diversity (~543 Ma), at which point all the known phyla had originated.

Though the Cambrian explosion saw the origin of all major body plans from flatworms to arthropods, the Ordovician period (~500 Ma) saw the diversification and the filling out of the lower-order taxonomic classes (Droser and Finnegan, 2003). The Ordovician radiation (or the Great Ordovician Biodiversification Event) gave rise to the flora and fauna that remained stable until the Permian-Triassic extinction (~251 Ma) (Munnecke et al., 2010). During this time, the first vertebrates with true bones arose (jawless fish, ~485 Ma), and primitive plants moved onto land (~434 Ma). By 420 Ma, insects began moving onto land, but the terrestrial biosphere would not have been recognizable until ca. 363 Ma, when amphibious tetrapods evolved adaptations that would allow for life out of the water.

The Permian-Triassic extinction (~251 Ma), the biggest in the history of Earth, saw the extinction of nearly 96% of marine and 70% of terrestrial multicellular species (Benton, 2005) and paved the way for the Mesozoic marine revolution, which fundamentally restructured marine ecosystems (Stanley, 1974; Vermeij, 1977). The dominant plant species during the Mesozoic were gymnosperms (cone-bearing, nonflowering plants), though angiosperms (flowering plants), the current dominant plant form on Earth, had evolved by the end of the Cretaceous (65 Ma) (Walker and Ward, 1992). The Mesozoic is known as “The Age of Reptiles” for a reason. The earliest dinosaurs evolved around 225 Ma and quickly became the dominant terrestrial predator. Mammals first appeared in the late Triassic (~200 Ma) but remained small and secondary until after the Cretaceous-Tertiary (also known as the “K-T boundary,” which is now properly referred to as the Cretaceous-Paleogene boundary) extinction event, 65 Ma (Luo, 2007). All terrestrial dinosaurs went extinct at the K-T boundary, though avian dinosaurs survived and evolved into modern birds during the Cenozoic (Hou et al., 1996). The secondary adaptive radiation of birds and mammals occurred in the Paleogene (a geological period from 66 to 23 Ma), at which time all modern groups of plants and animals are present.

FIG. 15. Comparison of the atmosphere of Earth (Rakhecha and Singh, 2009) with its neighbors Mars and Venus (William, 2011). Note the similarity in relative composition in the atmosphere of Venus and Mars versus the biologically cycled atmosphere of Earth. (Credit: Carie Frantz)
4.2.5. What causes mass extinctions, and what else controls biodiversity?  Biodiversity is the number of species that exist in a particular location at a particular time, although this type of quantification may become difficult when examining microbial communities where genetic exchange makes species boundaries more malleable. Biodiversity depends on two primary factors: rate of speciation (species generation) and rate of extinction. If mobility between regions is possible, regional biodiversity may also be affected by immigration and emigration. A region’s “carrying capacity” for diversity, or the maximum number of species it can support simultaneously, depends on the number of habitats and resources that can be utilized.

Examinations of biodiversity on geological timescales rely on Earth’s fossil record, which is overwhelmingly biased toward macroscopic life. Trends in the diversity of macroscopic fossils tend to indicate that their total biodiversity has increased over the course of geological time, as exemplified by the Sepkoski curve, which was compiled from an exhaustive survey of paleontological literature on marine organisms (Sepkoski, 1982, 1992, 2000). Studies in the past few years suggest the recent increase in biodiversity may be an effect of sampling bias (in that there is better preservation of younger fossils than older ones). However, the general trend of increasing biodiversity of macroscopic organisms with time still appears to hold (Alroy, 2008).

The primary factor that contributes to a species’ extinction is change in the environment beyond a threshold to which a species can adapt. Changes in environment can be either abiological or biological in origin. Abiological changes can include changes in nutrient content, temperature, atmospheric or ocean conditions, or any other perturbation of the geological and chemical conditions to which a species is well adapted. In addition to changes in the geological and chemical environment, biological interactions and evolution of novel traits can disrupt ecosystem organization, resulting in the extinction of species that cannot cope with the change.

Throughout Earth’s history, there are examples of fossilized species, once dominant in their ecosystem, going extinct due to environmental change or ecological pressure. When environmental changes become too extreme, the rate of extinctions can become high enough to qualify for the term “mass extinction.” This has occurred five times in Earth’s history. Although the exact causes and mechanisms remain unknown, most have been linked to rapid changes in climate (Table 3; Benton, 1995; Sepkoski, 2000; Bambach et al., 2004). Further, these data and their analyses are subject to preservation effects and the biases these introduce. For example, the recorded mass extinction events are specific to multicellular organisms that create well-preserved fossils.

Chapter 5. What Does Life on Earth Tell Us about Habitability?

5.1. What are the commonalities and common habitability requirements of known Earth life?

When considering whether an environment or planet is habitable, at a minimum scientists take into account the potential energy source (or sources), the ambient chemical composition and whether this composition is suitable to build biomass, and the availability of a solvent. Additionally, there are physical constraints upon life (e.g., temperature, pressure,

| Table 3. Earth’s Extinction Events (Credit to Lev Horodyskyj) |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| **Name**        | **Timing**      | **Phyla impacted**      | **Phyla extinct**      | **Causes**      | **References** |
| Ordovician-Silurian | 450–440 million years ago | Brachiopods, bryozoans, trilobites, conodonts, graptolites | Sudden short-lived glaciation event, sea level drop (related to nonvascular plant evolution?) | Finnegan et al., 2011; Lenton et al., 2012 |
| Late Devonian | 375–360 million years ago | Brachiopods, trilobites, rugose and tabulate corals | Unknown (related to vascular plant evolution?) | Buggisch, 1991; Algeo and Scheckler, 1998 |
| Permian-Triassic | 251 million years ago | Forams, radiolarians, bryozoans, brachiopods, bivalves, gastropods, ammonites, crinoids, insects, gymnosperms, seed ferns, amphibians, sauropsids, therapsids | Rugose and tabulate corals, blastoids, trilobites, eurypterids, acanthodian fish | Isozaki, 1997; Knoll et al., 1996; Kump et al., 2005; Renne et al., 1995; Erwin, 2000; Cao et al., 2009 |
| Triassic-Jurassic | 205 million years ago | Archosaurs, amphibians | Conodonts, therapsids | Unknown | Ward et al., 2001; Hesselbo et al., 2002 |
| Cretaceous-Tertiary | 65.5 million years ago | Foraminifera, coccolithophorids, corals | Ammonites, belemnites, rudist bivalves, mosasaurids, plesiosaurs, pterosaurs, dinosaurs | Extraterrestrial impact of asteroid | Molina, 1995; Toon et al., 1997; Alvarez et al., 1980 |
pH, salinity, radiation, metal toxicity; see Section 5.3 for environmental limits) such that the presence of these habitability requirements may be necessary but not sufficient for life (Jones and Lineweaver, 2010). The fundamental habitability requirements are certainly some of the commonalities of life on Earth. (Chapter 2 contains an inclusive list of common life requirements that collectively characterize “life” on Earth.)

- Energy: A source of energy is necessary for metabolism and to maintain the functions carried out by a cell. As previously stated in Section 4.2.1, life on Earth uses either solar energy or chemical energy sources that include both organic (carbon) compounds and inorganic elements and molecules.

- Common essential elements: All known Earth life uses the same main set of elements to form biomolecules (CHNOPS; see Sections 4.2.1 and 5.1.1) and shares a common carbon-based biochemistry. Carbon is an excellent element from which to build life, because carbon forms macromolecules with a variety of other elements and itself. Carbon forms primarily covalent bonds (compared to ionic bonds), and carbon-based molecules are very stable in the presence of water and oxygen (Schulze-Makuch and Irwin, 2008).

- Solvent: A solvent to host chemical interactions is considered a universal requirement for all life. With extensive evidence that all Earth life needs water specifically as that solvent (Mott et al., 2007), many astrobiologists consider water to be the most probable solvent for life on any planet.

Some authors have suggested a definition with fewer requirements; these definitions are still consistent with the characteristics of life on Earth but are less constraining. For example, Benner et al. (2004) named only the presence of both thermodynamic disequilibrium and a temperature consistent with chemical bonding as requirements for habitability. By decreasing the number of requirements to consider something as “living,” Benner et al. (2004) proposed to expand our paradigm of “life” beyond our Earth-centric view.

5.1.1. What is the chemical composition of Earth life?

The large diversity of extinct and extant life on Earth has all shared a common chemical composition. Biologists consider carbon, hydrogen, nitrogen, oxygen, sulfur, and phosphorus to be the six elements critical to sustain life (often referred to as CHNOPS), although at least 77 elements interact with life on Earth or are required for biological function (Committee on the Limits of Organic Life in Planetary Systems, 2007). These six elements are known to account for up to 97% of the mass in a cell (Escherichia coli; Wackett et al., 2004). All these elements have soluble or volatile forms that create element-specific biogeochemical cycles in the biosphere.

Carbon (C), hydrogen (H), nitrogen (N), and oxygen (O) are the most abundant elements in living cells. They are capable of forming a wide variety of chemical bonds, including hydrogen bonds, double and triple covalent bonds, and ionic bonds with each other, so they are useful elements with which to make organic matter. Carbon forms the backbone of molecules in organic matter and is found in almost all the molecules in organisms (except water). Nitrogen is a critical component of amino acids, the building blocks of proteins, and nucleic acids. Some microbes, called nitrogen fixers or diazotrophs, can use triple-bonded atmospheric nitrogen as a nitrogen source, and other microbes, called denitrifiers, release triple-bonded nitrogen gas as waste back into the atmosphere. Oxygen is also a key component of organic matter and water. On Earth, nearly all the double-bonded oxygen in the atmosphere is produced by photosynthesis. Many organisms (such as animals) use that atmospheric oxygen for high energy–yielding aerobic respiration. Hydrogen is found in almost all molecules in organisms as well, and its main importance to life comes from its ability to form hydrogen bonds. While hydrogen can form hydrogen bonds with carbon, hydrogen bonding with oxygen and nitrogen are most important for life’s basic processes. For instance, hydrogen bonding occurs between complementary base pairs (adenine-thymidine and guanine-cytosine) and thus brings together two strands of DNA (deoxyribonucleic acid). Hydrogen bonding also helps structure nascent proteins and partially gives water the property of high specific heat capacity (see Section 5.1.2 for more reasons why hydrogen in water is important to life). Phosphorus (P) is a component of important biological molecules, such as DNA, RNA (ribonucleic acid), membrane phospholipids (components of cell membranes), and ATP (the energy currency of the cell; see Section 2.2). Sulfur (S) is a critical component of two amino acids used by life.

Besides the CHNOPS elements, there are many additional elements that constitute living matter and interact with life (see Fig. 1 from Wackett et al., 2004, for a visual representation of which elements in the periodic table are essential to life). Some metals, such as calcium (Ca), sodium (Na), and potassium (K), are almost universally utilized by Earth-based life; these elements have critical roles in biochemical signaling pathways and help maintain a potential difference across biological membranes. In addition, calcium is used as a structural component in animal bones, marine invertebrates, and land-based snails (in the form of calcium carbonate in the latter cases). Many metals, such as iron (Fe) and magnesium (Mg), can be used as components of enzymes (biological catalysts that control chemical reactions inside cells) or cofactors (nonprotein chemical compounds necessary for enzymatic reactions) for enzymatic reactions across many taxa (groups of organisms). These metals can also be used in redox reactions for chemosynthesis or as electron donors for photosynthesis, as a source of energy for life (see Section 4.2.1 for more on these types of energy sources). Still, the use of some elements by biota can be taxon-specific; for instance, diatoms (a type of microbial phytoplankton) are dependent upon oceanic silicon to build their shells (called frustules).

We know that life (in particular microbial life) is capable of adapting its chemical composition to a huge range of environmental conditions. For example, it was once thought that all life used phospholipid membranes; however, research has shown that bacteria living in phosphorus-poor regions of the ocean are capable of substituting sulfolipids for phospholipids (Van Mooy et al., 2006). Further, there are many organisms that can substitute for specific metallic elements at the centers of metallo-enzymes. The possibility that life on Earth may have the ability to adapt to its environment by changing the elemental composition of some of its structures or fluids remains an intriguing and exciting question. However, to date, no organism has been found that has been universally accepted to be able to live without using the CHNOPS elements.
Wolfe-Simon *et al.* (2011) reported on a cultivated bacterium from an arsenic-rich lake. The paper’s authors claimed that this bacterium is capable of substituting small amounts of arsenic compounds for phosphorus in selected biomolecules, including nucleic acids and proteins. However, there has been much criticism of this paper (Benner, 2011; Borhani, 2011; Csabai and Szathmáry, 2011; Foster, 2011; Oehler, 2011; Redfield, 2011; Schoepf-Cothenet, 2011), specifically that the data are not strong enough to support the claims of arsenic-substituted DNA. Furthermore, two independent efforts have failed to replicate the main results in the original manuscript (Erb *et al.*, 2012; Reaves* et al.*, 2012).

In sum, life on Earth requires a good number of elements, and a select few (CHNOPS) in very high abundances. The chemical composition is similar among known Earth-based life, but many taxa have evolved to continuously adapt to their environment by slightly altering their chemical composition.

### 5.1.2. Does all life share a common chemical solvent?

All known Earth-based life uses water as a solvent, and some scientists suggest that Darwinian processes (e.g., evolution and adaptation) require water or other solvents similar to water (Committee on the Limits of Organic Life in Planetary Systems, 2007). A liquid solvent is considered necessary in order for the chemical reactions of life to occur and for life to maintain an appropriate temperature (e.g., Schulze-Makuch and Irwin, 2008). A biological solvent provides a localized region for chemical reactions, allows an organism to concentrate nutrients, and facilitates a system for waste products to be expelled from the cell. Water (H$_2$O), the solvent for all known Earth life, accounts for a large fraction of the body weight of most known living organisms. For instance, by mass, water makes up 40% of land plants, 60% of microbes, 70% of humans, and 95% of jellyfish (Wackett *et al.*, 2004; Schulze-Makuch and Irwin, 2008).

There are many likely reasons why life on Earth evolved to use water as a common solvent (Mottl *et al.*, 2007; Schulze-Makuch and Irwin, 2008), and this review will address the top three reasons. First, water is a very common molecule on Earth: hydrogen is the most common element in the Universe, and oxygen is the most common element on Earth (Mottl *et al.*, 2007). Second, water is a polar molecule with hydrogen bonding capability (see the work of Mottl *et al.*, 2007, for an explanation of water properties resulting from its dipole structure and the work of Pohorille and Pratt, 2012, for more discussion of water’s properties). The polarity of water allows it to dissolve most other polar molecules, including salts for metabolic reactions. Third, water provides electrical insulation from surroundings (a high dielectric constant) as well as thermal stability and insulation (a high specific heat capacity). This allows organisms with high water content and/or immersed in water to be somewhat shielded from environmental fluctuations (Schulze-Makuch and Irwin, 2008). The combination of these three characteristics and several more qualities (for a complete list, see the works of Mottl *et al.*, 2007, and Schulze-Makuch and Irwin, 2008) make water a useful solvent for life.

Taking this large importance of water to Earth-based life into account, astrobiologists interested in detecting life beyond Earth give strong preference to sites with the potential for present water or where water may have existed in the past. This “follow the water” motto has been used to help determine sites of astrobiological interest on other bodies in the Solar System and beyond (Committee on the Limits of Organic Life in Planetary Systems, 2007; Chapter 6). However, water is not the sole environmental requirement for habitability, as discussed above, and the presence of liquid water does not necessarily indicate that life will be found. Jones and Lineweaver (2010) estimated that 88% of Earth’s volume where free liquid water exists may not be inhabited by life, possibly because of constraints from temperature, pressure, nutrient availability, pore space, and lack of energy sources. With this in mind, other guidelines are also used in conjunction with “follow the water” to determine whether extraterrestrial sites are of potential astrobiological interest. Another approach is to identify sites with an abundance of elements needed for life; for instance, past and present organic carbon is a priority for Mars astrobiology science (Committee on an Astrobiology Strategy for the Exploration of Mars, 2007). Because of the universal need for organisms to use energy sources from the environment, Hoehler *et al.* (2007) suggested to “follow the energy.”

### 5.1.3. What do we know about the minimum energy requirements of Earth life?

#### 5.1.3.1. The simple circuit analogy

All life requires some amount of energy. This energy, as was discussed in Section 4.2, is obtained by generating electron flow created through redox reactions of environmental compounds. Even in photosynthesis, where sunlight is the energy source, redox reactions are used as a means of capturing light energy and converting light energy into chemical energy that the cell can use (see Section 4.2). The transfer of energy via electrons is similar to the way in which a light bulb receives energy from a battery in a simple circuit (Fig. 16). The energy in the simple circuit must be delivered to the bulb at a minimum amount (voltage) and a minimum rate (power) in order to generate light. Life has the same constraints in order to maintain metabolism, and like power and voltage in an electric circuit, these constraints can be quantified.

Much like the voltage in a simple circuit, there is a minimum amount of energy required to drive metabolism. This is described by the quantity called Gibbs free energy, or $\Delta G$, which refers to the energy available in a system to do work. This threshold is generally accepted to be about $-20$ kJ/mol in actively growing microbial populations, which is approximately $1/3$ of the $-70$ kJ/mol required to build 1 mol of ATP, the molecule that stores energy in all known Earth life (Schink, 1997). Studies have shown that this minimum amount of free energy can be as low as $-9$ kJ/mol in energy-limited populations where organisms may be more efficient in their energy utilization (Hoehler, 2004; Schink and Stams, 2013).

Maintenance energy (ME) is defined as the minimum rate at which energy must be delivered to an organism to maintain cellular structure and function (Pirt, 1965, 1982; Neijssel and Tempest, 1976). In the analogy of a cell as a simple circuit, the ME represents the power of the organism and can be theoretically calculated similarly to the calculation of power in the circuit (Fig. 17). This calculation of the ME demands a minimum rate of substrate mass transport into an organism, similar to the current term ($I$) in the electrical power calculation, which quantifies the rate of electron flow. This translates into substrate utilization flux ($J$) of the limiting reagent in the redox reaction.
5.1.3.2. Quantitative habitability and energy balance. Environments that offer abundant and accessible energy resources are high-priority exploration targets for the astrobiology community (Link et al., 2005; Hoehler et al., 2007; MEPAG, 2010). As described in Section 5.3.1, life can be found in just about any environment on Earth that contains liquid water, a sufficient energy source, and the right elements. However, the capability of any given environment to support a microbial population is derived from the balance between energy supply, energy demand, and the effects of environmental conditions, such as temperature, pH, and salinity (Hoehler, 2007). Further, if we can quantify both supply and demand, the balance of the two becomes a potential quantitative tool for evaluating habitable zones for exploration.

In this conceptual model, the generation of energy supply is mediated by the thermodynamics and kinetics of available redox reactions, and much work has been done to understand the impact of other environmental constraints (Amend et al., 2003; Shock et al., 2010). Energy demand, on the other hand, is slightly more complex, as many environmental effects on ME are not well constrained. This lack of knowledge presents limitations for developing accurate quantitative models; thus ME is still empirically derived (Shock and Holland, 2007; Smith and Shock, 2007). Therefore, a “follow the energy” approach is more difficult to translate into measurements and would likely require a broader suite of measurements than a “follow the water” approach (where one “simply” searches for the presence of water). Additional work could improve the ability of this methodology to be implemented in the search for life on other planets, and a “follow the energy” strategy is gaining momentum in the community due to its potential for the determination of habitable zones on Earth and elsewhere.

5.2. What are potential exceptions to the commonalities shared by life on Earth?

All known life on Earth shares a common ancestry (see Section 4.1). Thus, all life may have inherited its biochemistry from that common ancestor, and the study of currently known Earth life may not provide a complete view of the full range of biochemistries that might support life in the Universe. Furthermore, there are several universalities of known life that are not necessarily predicated by prebiotic chemistry. For example, the amino acids used by life show little correlation with abiotic synthesis of amino acids (Ward and Benner, 2007), and only four bases are used in nucleic acids.
when others are also possible (Benner and Sismour, 2005; Yang et al., 2006). Thus, some astrobiologists consider possibilities for life with different biochemistries and even speculate about the possibility of an as-yet-undiscovered Earth life-form possessing an expanded or entirely different biochemistry (Davies et al., 2009). The possibility of such a “shadow biosphere” existing has implications for the discussion in this entire section, as it highlights that our expectations for life beyond Earth are based on all known Earth life. Any undetected Earth-based life could have significantly different characteristics, and our lack of knowledge of these organisms would significantly bias our search for life beyond Earth.

5.2.1. Could non-carbon-based life exist? All known biochemistry is carbon-based, despite the fact that Earth is overwhelmingly silicon-rich; the silicon-to-carbon ratio in Earth’s crust is about 925:1 (Haynes, 2012). Carbon is able to bond with four other atoms (as is silicon), and the energy required to make or break bonds is at just the right level for creating molecules that are stable, but not too stable. This means that carbon remains reactive at the range of temperatures typically found on Earth. In other planetary environments, however, temperatures may be too low or high, with carbon bonds becoming too stable or unstable, respectively.

Silicon-based life has been proposed as a potential alternative to carbon-based life because silicon has several chemical properties similar to carbon. Silicon is part of the same group in the periodic table (column 14) and, like carbon, is able to form stable polymers with itself (silanes); with carbon, nitrogen, phosphorus, and oxygen (silicones); as well as with many metals (Bains, 2004). Silicon polymers can become large enough to form large molecular backbones and ring structures analogous to those in the molecules that encode information in known organisms. Therefore, silicon can match carbon in terms of chemical flexibility and functionality, albeit not by using the same chemistry to achieve the same goal (Bains, 2004). One issue, however, is that Si-O bonds are generally too stable, while Si-Si and Si-H bonds are too unstable (for life that uses water as a solvent).

Biogenic silica is extensively utilized by life on Earth in a wide variety of forms and by a wide variety of organisms (Birchall, 1995). In many plants and algae, for example, silicon dioxide plays an important structural and defensive role in their cell walls, which can account for up to 10% of their dry weight. The prevalence of silicon in these organisms suggests that extraterrestrial life-forms could at least have similar silicon-based external structures.

Although some have argued that silicon is poorly suited for life (Pace, 2001), there are others who make the point that different factors will apply in extraterrestrial environments with chemistry that is very different from Earth. For example, cold methane oceans such as those found on Titan (Larson and McKay, 2013) could be more favorable for silicon-based life. In such environments, the greater reactivity of silicon-based chemicals could be an advantage, enabling living processes to occur (Bains, 2004).

5.2.2. Could solvents other than water be used? All known life requires water. As set out in Section 5.1.2, water confers many advantages as a biosolvent, as follows: (1) water has a high heat capacity (the amount of heat required to change the temperature of a substance by a given amount) and effectively stabilizes the environment against rapid temperature swings; (2) water is liquid over a large temperature range; and (3) water can dissolve a wide variety of compounds. A water solvent is necessary for membranes and macromolecules to form the correct three-dimensional structure (Alpert, 2006). Without water, essential macromolecules aggregate, and internal ultrastructure is lost, although some species have special adaptations that allow them to rebound from severe desiccation (Potts, 1994, 2001; Alpert, 2006). This lack of water is one of the most difficult conditions for life to face (Alpert, 2005).

However, some of the properties of water present challenges for life. On a global scale, feedbacks in which water is involved can lead to the instability of liquid water. See Chapter 6 for a discussion of these feedbacks and how they impact our search for habitable environments beyond the Solar System. Another disadvantage is that the chemical reactivity of water leads to degradation of essential bio-polymers such as DNA and RNA, which must be continuously repaired in Earth-based life (Ward and Benner, 2007). This damage leads to significant challenges in synthesizing these polymers abiotically under aqueous conditions (Engelhart and Hud, 2010). However, the fact that all known life on Earth uses water as a solvent demonstrates that any disadvantages of water are not insuperable.

The question of whether alternative solvents to water might be used, and in what environments, is still debated within the astrobiology community, although the view of many is still that liquid water is an essential condition for life (Bains, 2004; Benner et al., 2004; Schulze-Makuch and Irwin, 2006; Ward and Benner, 2007). Alternative solvents include sulfuric acid, ammonia, methane, ethane, nitrogen, each of which would be stable at the surface for different star-planet orbiting distances.

Ammonia was one of the earliest alternative solvents proposed (Haldane, 1954). Ammonia is of particular interest, as its freezing point is much lower than the freezing point of water (Elliott, 1924); so if it can be used as a solvent, it could potentially allow for habitability at much lower temperatures. Liquid ammonia shows some striking chemical similarities with water, with the added advantage of dissolving most organics as well as, or better than, water (Bains, 2004). Life in an ammonia or water-ammonia mixture would require novel compartmentalization, since lipid-based membranes (a major component of cell walls in Earth life) dissolve in ammonia. Additionally, whereas Earth life exploits compounds with oxygen doubly bonded to carbon, ammonia-based life might instead implement nitrogen doubly bonded to carbon (Benner et al., 2004). Other solvents suggested for extreme cold environments include dinitrogen (N2) and supercritical dihydrogen-helium mixtures (Bains, 2004; Benner et al., 2004).

On the other temperature extreme, concentrated sulfuric acid has been considered a solvent that might support life at extremely high temperatures (Schulze-Makuch et al., 2004). Additionally, the idea of methane as a potential solvent for life is gaining interest since the discovery of methane lakes on Titan (Stofan et al., 2007). However, life that utilizes methane in place of water is difficult to envision in reference to Earth-based life, given that methane is a nonpolar solvent with chemical properties widely different from water. Other, less abundant solvents that have been considered as potentially life-supporting include methyl alcohol, hydrogen sulfide, hydrogen fluoride, hydrogen cyanide, and hydrogen chloride (Bains, 2004).
5.2.3. Could other energy sources be utilized? A much larger variety of potential energy sources than just solar or chemical energy are available both on Earth and other planetary bodies. In particular, known life can respond to a wide range of energy sources not currently utilized in biological systems, including thermal energy, pressure, movement of water and air, gravity, and electromagnetic fields (Muller and Schulze-Makuch, 2006). Could any of these sources be used in place of life’s current inventory of energy sources? Thermal energy could potentially drive life processes, given the ubiquity of thermal energy and thermal gradients in the Universe (Muller and Schulze-Makuch, 2006). Alternative life is driven by the movement of air rather than living in an aqueous environment. Such a hypothesis was first proposed by Sagan and Salpeter (1976) for potential life-forms in Jupiter’s atmosphere where ecological niches occur that could accommodate distinct gas-filled organisms such as “sinkers,” “floaters,” and “hunters.” Even more extreme varieties of life have been postulated, which utilize gravitational energy (Drake, 1973) or even nuclear decay (Adam, 2007). However, the fact that none of these sources of energy are used by life on Earth suggests that they are all, at best, significantly less effective (based on known biochemical mechanisms for harnessing energy) than redox chemistry and that redox chemistry is the most likely energy source to be used by any life beyond Earth.

5.2.4. Could there be a different genetic code? The genetic code of life defines a mapping between DNA sequences and the sequence of amino acids in proteins. All life on Earth shares a common genetic code that has only minor variations between different groups of organisms. There could, in principle, be different genetic codes in alternative biochemistries. It is an open question as to whether the genetic code of modern life was fixed by the chemical affinities between codons and amino acids (Yarus, 2000) or the result of an evolutionary optimization process (Freeland and Hurst, 1998; Knight et al., 2001). Life incorporating different amino acids than those used on Earth would necessarily utilize genetics with a different genetic code (i.e., a different mapping between nucleotide codons and translated amino acids). Alternative codes could also be used even if the same amino acids were used. Experiments have demonstrated that the machinery of known life can be manipulated to incorporate many nonbiological amino acids into proteins during translation (Bain et al., 1989; Noren et al., 1989; Chin et al., 2003). Another alternative genetic system could also have a genetic polymer (i.e., DNA-like molecule) with different bases than the four used on Earth or with a different chemical backbone (Geyer et al., 2003; Benner et al., 2004). Artificial genetic systems of this expanded set have been demonstrated to sustain replication for up to 20 generations (Sismour and Benner, 2005). More extreme hypothetical cases include life-forms that do not use polymers to encode genetics, such as quantum-based life scenarios (Davies, 2004).

5.3. What are the environmental limits for life as we know it?

5.3.1. Which environments are considered to be extreme on Earth? Astrobiologists use the universal requirements for life as far as they are known (see Section 5.1) to constrain possible extraterrestrial habitats. By examining the empirical and theoretical limits to life on Earth, it is possible to narrow down the search to just habitats that Earth-like life could currently inhabit or could have inhabited at some point in the past. Over the past several decades, scientists have discovered microorganisms in extreme environments, also known as extremophiles, which are continually expanding our knowledge of the limits of tolerance. For this reason, the absolute physical and chemical limits of life on Earth are not known. The currently measured extremes of the physical and chemical parameters of habitability on Earth are shown in Table 4. It is important to note that the values shown are only the limits for metabolically active organisms and that dormant life (such as a spore) is often

### Table 4. The Known Limits of Life on Earth

<table>
<thead>
<tr>
<th>Factor</th>
<th>Extremophile</th>
<th>Earth environments</th>
<th>Limit</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>High temperature</td>
<td>Thermophile</td>
<td>Submarine hydrothermal vents, terrestrial hot springs, deep subterranean environments</td>
<td>121–122°C</td>
<td>1–3</td>
</tr>
<tr>
<td></td>
<td>or hyperthermophile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low temperature</td>
<td>Psychrophile</td>
<td>Ice, permafrost, Antarctic dry valleys</td>
<td>−20°C</td>
<td>4–5</td>
</tr>
<tr>
<td>High pressure</td>
<td>Barophile or piezophile</td>
<td>Subseafloor, deep subterranean environments such as the Mariana Trench</td>
<td>&gt;1300 atm</td>
<td>6–8</td>
</tr>
<tr>
<td>High pH</td>
<td>Alkaliphile</td>
<td>Soda lakes</td>
<td>11.3–11.4</td>
<td>9–11</td>
</tr>
<tr>
<td>Low pH</td>
<td>Acidiphile</td>
<td>Acid mine drainage, hot springs</td>
<td>−0.06</td>
<td>12–14</td>
</tr>
<tr>
<td>High salinity</td>
<td>Halophile</td>
<td>Brines, evaporite deposits</td>
<td>NaCl ≈ 5 M</td>
<td>15–17</td>
</tr>
<tr>
<td>Desiccation</td>
<td>Xerophile</td>
<td>Antarctic dry valleys, deserts</td>
<td>&lt;10% relative humidity</td>
<td>18–20</td>
</tr>
<tr>
<td>Ionizing radiation</td>
<td>Radiosensitive</td>
<td>High-radiation environments (from radioactive waste, nuclear reactors, cosmic rays, gamma rays, X-rays)</td>
<td>&gt;6000 Gy</td>
<td>21–23</td>
</tr>
<tr>
<td>UV radiation</td>
<td>Radiosensitive</td>
<td>High deserts, upper atmosphere</td>
<td>5000 J/m²</td>
<td>21, 24</td>
</tr>
<tr>
<td>Toxic element concentrations</td>
<td>—</td>
<td>Acid mine drainage, hot springs</td>
<td>Depends on element</td>
<td>25–26</td>
</tr>
</tbody>
</table>

able to tolerate even more extreme conditions (Rothschild and Mancinelli, 2001; Cavicchioli, 2002; Jönsson et al., 2008). These are two end-members on the spectrum of environments, with environments in which life can live and grow on one side and environments in which life can survive (to grow at a later time or different place) on the other side. Since there is still much to learn about extremophiles and their adaptations, it is likely that the defined limits for the survival of Earth-like life will continue to expand.

Several of the limits shown in Table 4 are important for habitability because they control the availability of liquid water. For example, temperature and pressure are key constraints on the boiling and melting points of water. Other environmental conditions are important because they affect the ability of atoms to form and maintain chemical bonds; this is why extremely high temperatures or high levels of radiation are not considered habitable environments. Temperature and pH control the stability of organic molecules and the kinetics of chemical reactions. Salinity also controls the amount of liquid water available to the cell, causing osmotic stress at the extremes.

In many environments, the extreme factors are compounded, making it necessary for extremophiles to develop multiple adaptation strategies. For example, high-pressure environments are often associated with high temperatures, such as in submarine hydrothermal vents or the deep continental subsurface (e.g., Takai et al., 2008), and low pH often causes an excess of heavy metals (e.g., Robbins et al., 2000). There are many different types of adaptations extremophiles have developed that enable them to tolerate these environmental extremes. Tolerance to high and low temperatures involves the production of specific enzymes to maintain chemical reactions and changes to the lipid composition of membranes to maintain membrane fluidity (e.g., Feller and Gerday, 2003). For chemical extremes, such as high and low pH, high salinity, and high concentrations of heavy metals, many of the adaptations involve enhancing cell walls, altering the internal chemistry, and producing special transport proteins to control entry and exit of molecules (e.g., Baker-Austin and Dopson, 2007). Other common adaptations improve DNA repair mechanisms; these are particularly important for organisms exposed to high doses of radiation or subjected to severe desiccation (e.g., Potts, 1994; Mattimore and Battista, 1996). These processes require additional energy. The assessment of an environment’s habitability, therefore, needs to take into account not only the environmental conditions but also whether there are sufficient energy sources available to enable cells to cope with the conditions (Hoehler, 2004).

5.3.2. Can microbes survive in space? Space is perhaps the most extreme environment where some forms of life can still maintain their shape, structure, and potential to live, even though growth or multiplication has not been observed and is not thought to be likely. The physical parameters that life must adapt to in space include low pressure (vacuum), extreme temperature variations (depending on their orientation relative to the Sun), high levels of different types of radiation, and microgravity.

The effects of these parameters have been extensively tested on different forms of life since the 1930s: in simulation experiments (reviewed by Olsson-Francis and Cockell, 2010), by flight missions in Earth orbit, and also outside Earth’s orbit during NASA’s Apollo program (reviewed by Horneck et al., 2010).

The majority of the experiments used bacteria, archaea, or viruses as biological models in experimental or theoretical studies (Mileikowsky et al., 2000; Horneck et al., 2010). Among the important findings, spores of Bacillus subtilis survived the effects of the space environment for 6 years on board the NASA Long Duration Exposure Facility (LDEF), when protected from solar UV radiation (Horneck et al., 1994). Lichens survived in space after exposure to cosmic radiation and UV doses several times higher than those expected at sea level for the same 2-week period of exposure (Sancho et al., 2007). The lichen upper cortex (outer rind) seems to provide adequate protection against solar radiation. After extreme dehydration induced by high vacuum, the lichens were able to recover their full metabolic activity within 24 h (Sancho et al., 2007). In ground-based laboratory simulation experiments, Deinococcus radiodurans survived exposures to UV and solar low-energy charged particle radiation at doses comparable to those expected during a short trajectory between Mars and Earth (Paulino-Lima et al., 2010, 2011).

The study of microbial survival in space is important to understand basic biological mechanisms, such as the role of gravity on living beings; the effects of the radiation field in space; the survival factors in the upper boundary of Earth’s biosphere; and the likelihood of lithopanspermia, the interplanetary transport of microorganisms via meteorites (Horneck et al., 2010). This information is also crucial for spacecraft applications, such as the use of microorganisms in life-support systems; the monitoring, characterization, and control of spacecraft microflora; and associated crew health concerns (Horneck et al., 2010). It is also important to the development of planetary protection guidelines for robotic space missions, in particular those that seek to return samples of other objects to Earth (see Chapter 8).

5.3.3. How do we define habitable zones? 5.3.3.1. The habitable zone around the Sun. Because life on Earth is our only current example, many studies try to define what makes Earth’s location in the Solar System suitable to life. All life on Earth requires liquid water (see Section 5.1.2). Its presence is generally considered the most important characteristic of a “habitable planet,” although other possible solvents for life have been proposed, as discussed in Section 5.2.2. The circumstellar “habitable zone” is defined as the region around a star where the long-term presence of water on the surface of a planetary body with an atmosphere cannot be ruled out (e.g., Kasting et al., 1993, and Abe, 1993). Because the habitable zone is based on a definition of surface (rather than underground) liquid water, it significantly under-represents regions of space where liquid water—and life—may exist, though not at the surface, as explained in Section 5.3.3.4. However, the habitable zone is nevertheless considered by most astrobiologists to be a useful concept in beginning our search for life beyond the Solar System.

The location of the habitable zone is determined by a number of factors, including the composition and thickness of the planet’s atmosphere, biogeochemical cycles, and changes
in stellar output. For a star like the Sun, the current habitable zone has been conservatively estimated to extend from 0.95 to 1.37 astronomical units (AU) (e.g., Kasting et al., 1993). An astronomical unit is defined as the distance between Earth and the Sun, approximately 150 million kilometers. The new Kopparapu et al. (2013) habitable zone model recalculates the classic Kasting et al. (1993) habitable zone with updated model inputs, and its habitable zone extends from 0.99 to 1.70 AU. In this updated conservative estimation, Earth is very near the inner edge of the current habitable zone. However, this study did not include the effects of clouds or circulation of moisture, either of which could stabilize a planet at the inner edge of the habitable zone. Other studies have included these effects by studying the problem with global climate models. This pushes the inner edge of the habitable zone closer to the star, to 0.95 AU in one study (Leconte et al., 2013) and to 0.93 AU in another (Wolf and Toon, 2014).

The Sun, like all stars, gets brighter over its lifetime. As it fuses hydrogen (H) into helium (He) in its core, the number of nuclei decreases (four hydrogen nuclei become one helium nucleus). This change in the number of nuclei results in an increase in the temperature (and therefore the luminosity) of the Sun’s core required to maintain the equilibrium between the inward force of gravity and the outward thermal pressure (see Chapter 3 for more on stellar evolution). As a result of this stellar temperature evolution, the inner and outer habitable zone boundaries are not static; they are pushed outward as the Sun evolves. Thus, the “continuously habitable zone” (CHZ) is defined as the region where conditions for liquid water are present over geological timescales (defined here as the age of Earth). For the Solar System, the 4.6-billion-year CHZ has been estimated to extend from 0.95 to 1.15 AU (Kasting et al., 1993).

Mars orbits outside the outer edge of the solar habitable zone as defined by Kasting et al. (1993). Although the water on or near the martian surface today is frozen, there is evidence that the Red Planet had liquid water in the past (see Section 6.1.2). Selsis et al. (2007) defined the CHZ such that it includes the orbit of Mars in response to empirical geological evidence of liquid water, placing its outer edge at 1.77–2.4 AU. Such observations of liquid water set the outer edge of the “empirical habitable zone.”

Another alternative habitable zone definition suggested by Abe et al. (2011) is based on a habitable, though “dry,” planet with significantly less water than Earth. In this case, the habitable zone is 3 times wider than the above definition where a “wet” Earth-like planet is the standard for habitability. Pierrehumbert and Gaidos (2011) suggested a H$_2$He greenhouse habitable zone for hypothetical planets with very thick atmospheres (these atmospheres may be accreted during planet formation). The low molecular weight of hydrogen makes it susceptible to loss to space when stellar charged particles and UV radiation heat the upper atmosphere, but Pierrehumbert and Gaidos (2011) suggested that thick hydrogen atmospheres may be stable on Earth-sized planets greater than 2 AU away from solar-type stars. A 40 bar H$_2$He atmosphere can push the outer limit of the habitable zone out to 10 AU for a 3-Earth-mass planet around a solar-type star. However, only the stability of surface oceans has been considered for such planets; other criteria for habitability have not yet been considered. These habitable zones are summarized in Table 5.

### Table 5. Different Definitions of Habitable Zone Boundaries around Solar-Type Stars

<table>
<thead>
<tr>
<th>Habitable zone</th>
<th>Inner edge (AU)</th>
<th>Outer edge (AU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical</td>
<td>0.95</td>
<td>1.77–2.4</td>
</tr>
<tr>
<td>“Standard” model</td>
<td>0.95</td>
<td>1.37</td>
</tr>
<tr>
<td>“Standard” model CHZ</td>
<td>0.95</td>
<td>1.15</td>
</tr>
<tr>
<td>“Kopparapu” model</td>
<td>0.99</td>
<td>1.70</td>
</tr>
<tr>
<td>“Dry” model</td>
<td>0.72–0.95</td>
<td>1.37–1.5</td>
</tr>
<tr>
<td>H$_2$ greenhouse</td>
<td>~2</td>
<td>&lt;10</td>
</tr>
<tr>
<td>“Leconte” greenhouse</td>
<td>0.95</td>
<td>no value</td>
</tr>
<tr>
<td>“Wolf and Toon” greenhouse</td>
<td>0.93</td>
<td>no value</td>
</tr>
</tbody>
</table>

The inner edge provided for the empirical habitable zone (Selsis et al., 2007) and the standard model (Kasting et al., 1993; Kopparapu et al., 2013) is based on the location of significant water loss. The outer edge of the standard model is where CO$_2$ cloud condensation occurs. The empirical habitable zone’s outer edge is defined to include Mars at 1.5 AU (4 billion years ago) because there are geological signs of past water. The inner edge of the “dry” habitable zone is between Earth and Venus; its outer edge is between Earth and Mars (Abe et al., 2011). The H$_2$ greenhouse habitable zone is for a 3-Earth-mass planet with a 40 bar H$_2$ atmosphere (Pierrehumbert and Gaidos, 2011).

5.3.3.2. Habitable zones around other stars. The location of the habitable zone depends on how bright the star is (Kopparapu et al., 2013); for brighter stars, the habitable zone is shifted outward because planets must be farther away from these stars to maintain temperatures necessary for liquid water, and for dimmer stars, the habitable zone is shifted inward (Fig. 18). A star’s lifespan is inversely related to its luminosity. Therefore, dimmer stars live longer than bright stars and evolve more slowly. Due to this slower change in brightness, relatively large portions of the habitable zones around dimmer stars stay habitable for long periods of time. But that length of habitability comes at a cost; for much of the lifetime of these stars, they will be “highly active” and emit high-energy radiation that could strip away the atmospheres of any planets in the habitable zone (Luger and Barnes, 2015). Another consideration relevant to low-mass stars is how close a planet can come to the star without becoming tidally locked (known as the star’s tidal-locking radius). One concern with this has been that the “dark side” of the planet will be cold enough for the atmosphere to freeze out. However, sufficient orbital eccentricity (eccentricity measures orbital shape; a perfectly circular orbit has zero eccentricity, with higher numbers reflecting more flattened ellipses) and interactions with other planets may be enough to prevent such a fate (e.g., Barnes et al., 2008). Furthermore, the atmospheres of synchronously rotating planets may not freeze out if atmospheric circulation to the dark side of the planet transfers sufficient heat to warm (Edson et al., 2011, 2012). For these reasons, the aforementioned potential for low-mass stars to strip away atmospheres is currently thought to be a bigger barrier to habitability than tidal locking.

5.3.3.3. The galactic habitable zone. The galactic habitable zone (GHZ) concept was first introduced by Gonzalez et al. (2001). This concept describes the region in a galaxy where the origin and evolution of life is more favorable than
elsewhere. It is based on a set of properties including the locations of suitable stellar hosts, adequate distance from damaging radiation environments, and concentrations above a minimum threshold of the elements required to build rocky planets and the molecules needed for life (Gonzalez, 2005). The same considerations that apply to our galaxy apply to other galaxies as well, and it should be possible to define habitable zones for galaxies throughout the Universe.

Some doubt has been thrown onto the idea of a GHZ with the discovery of radial migration of stars through the Galaxy (Roskar et al., 2008). Stars that experience radial migration will encounter regions of the Galaxy with varied conditions. Therefore, it becomes difficult to define regions where life is more or less likely to exist if stars carry their planets in and out of these regions on short timescales.

5.3.3.4. Limitations of the habitable zone concept. When discussing the circumstellar habitable zone, it is important to keep in mind that it is a practical concept designed to focus our search for global, surface biospheres on planets around other stars. It is not a theoretical concept that accounts for all the areas in a planetary system in which liquid water and, therefore, life can exist. In this respect, there have been proposals of habitable zones for moons around giant planets, based on the proper orbiting distance to generate sufficient tidal energy to maintain liquid water in the subsurfaces of those moons. The “traditional” habitable zone, for planets, refers only to liquid water on the surface, but we know from Earth that life can also thrive underground. Liquid water can exist under the surfaces of planets and moons that are well outside the habitable zone. The habitable zone, therefore, significantly underrepresents the regions of space where life may be possible.

In the Solar System, all the worlds (besides Earth) that are considered potentially habitable lie outside the solar habitable zone. These worlds include Mars (depending on the outer edge of the habitable zone), Europa, Enceladus, and Titan (see Chapter 6). Liquid water may be possible in the subsurface of present-day Mars (e.g., Boston et al., 1992; McEwen et al., 2011; Ojha et al., 2015). Europa and Enceladus, moons of Jupiter and Saturn respectively, are thought to have subsurface reservoirs of liquid water (Kivelson et al., 2000; Collins and Goodman, 2007). Titan, with lakes of liquid ethane and methane, invites speculation on the possibility of life existing without liquid water at all (McKay and Smith, 2005) and could even harbor a “double biosphere” with an ethane/methane-solvent biosphere at the surface and water-solvent biosphere in the deep subsurface (Iess et al., 2012). In addition, there are other outer Solar System bodies that may contain subsurface oceans or liquid water (see Section 6.2).

### Chapter 6. What Is Known about Potentially Habitable Worlds beyond Earth?

Three necessary conditions for the survival of life as we know it are the availability of liquid water, raw ingredients...
6.1. What do we know about the habitability of the inner Solar System?

The inner Solar System consists of Mercury, Venus, Earth, Mars, and respective moons. Due to their lack of liquid water, Earth’s moon and the planet Mercury are not considered to be habitable or ever to have been habitable and are therefore not considered further here.

6.1.1. Is Venus habitable now? Was Venus ever habitable? We see Venus as the “Morning Star” and “Evening Star,” the brightest object in the night sky other than the Moon. Venus appears bright not only because of its close proximity but also due to the presence of a global cloud cover that very efficiently reflects sunlight. Up until the mid-20th century, the clouds were thought to be water, and Venus seemed to be the most likely candidate for another habitable planet in the Solar System. However, modern work has shown that the clouds consist primarily of sulfuric acid (Esposito et al., 1983), that Venus’ surface and atmosphere are devoid of water, and that the surface is around 740 K (872°C; Hunten, 2007). Venus is not habitable today but might have been early in its history.

Venus has a 92 bar atmosphere (1 bar being approximately the pressure of Earth’s atmosphere at sea level), consisting of 96.5% CO2 and 3.5% N2 (Fegley, 2003). Carbon dioxide, the dominant constituent, is an effective greenhouse gas. The greenhouse forcing from CO2, coupled with Venus’ proximity to the Sun, gives Venus an exceptionally hot surface, which excludes the possibility of life and has made in situ measurements extremely difficult (see Chapter 5 for a discussion of temperature and pressure limits on life).

Comparison between Venus and Earth suggests that Venus’ volatiles (elements or molecules that are relatively easily evaporated or sublimated) reside predominantly in its atmosphere. After scaling for the difference in planetary mass, Venus has 10,000 times the amount of carbon dioxide in its atmosphere than Earth’s atmosphere and oceans have combined, and 3 times the amount of nitrogen. However, most of Earth’s carbon (Zhang and Zindler, 1993) and at least half of Earth’s nitrogen (Goldblatt et al., 2009) are stored in rocks; once these geological reservoirs are accounted for, Venus and Earth have roughly the same amount of both elements.

Despite similar inventories of carbon and nitrogen, Venus is severely depleted in water relative to Earth. Contemporary Venus has no oceans, and water vapor is present in the venusian atmosphere at a concentration of only ~30 ppmv (parts per million by volume; Meadows and Crisp, 1996). Earth, by contrast, has deep oceans that would give a 260 bar atmosphere if evaporated completely. The difference in water inventories can be explained by the evolution of Venus’ atmosphere, with a “runaway greenhouse” process (described below) that caused the evaporation of any early oceans that may have existed (Hashimoto et al., 2008) or prevented condensation of oceans with subsequent loss of water by hydrogen escape (also see Section 5.2.2).

As Venus is closer to the Sun than Earth, the rate at which it absorbed solar energy would have exceeded the rate at which a moist atmosphere could radiate to space in the thermal infrared (IR) region. Venus would then have experienced runaway warming, which would have evaporated any early ocean (liquid water can only exist below the critical point, 374°C or 647 K), and Venus’ atmosphere would then have been mostly steam. Thermal balance in a steam atmosphere would be reached when the surface temperature is around 1700 K (1400°C); the surface itself does not emit radiation directly to space, but the emitting level of the atmosphere is hot enough to emit radiation in a mid-IR window where water vapor is a weaker greenhouse gas (Goldblatt et al., 2013).

In the upper atmosphere, water molecules would have been rapidly dissociated by photolysis (broken up by energy from incoming sunlight), and the resulting hydrogen would have escaped to space. The evidence for extensive hydrogen escape comes from measurements of the hydrogen isotopes in Venus’ atmosphere; deuterium (²H, also commonly denoted by D) is enriched 100-fold relative to Earth. Lighter ¹H atoms overcome gravity and escape more easily than ²H atoms, so the excess of ²H supports the hypothesis of hydrogen loss by escape and a previously wet Venus (Donahue et al., 1982). The rate of hydrogen escape would have been controlled by the amount of energy available; loss of an ocean the size of Earth’s present ocean would have taken a few hundred million years (Watson et al., 1981).

Transition from a runaway greenhouse to the bulk composition and climate of Venus’ atmosphere today requires the loss of water, addition of carbon dioxide, and development of thick cloud layers. Hydrogen escape evidently contributed to water loss, though hydration of surface minerals may also have contributed. Loss of water mass makes the atmosphere thinner and consequently cools the surface. If carbon dioxide was locked up in carbonates on a hypothetical temperate early Venus, these would have thermally decomposed during the runaway greenhouse phase. The thick CO₂ atmosphere now contributes to warming, given the absence of water vapor. Venus today is, in fact, much cooler than it could be because of the presence of high-altitude sulfuric acid clouds, which reflect most incoming sunlight.

It is unknown whether water ever condensed to form an ocean on Venus. One scenario would have an ocean condensing after planetary accretion (Hashimoto et al., 2008), giving a habitable period on early Venus. However, a runaway greenhouse may have followed directly from the magma ocean stage of planetary accretion, meaning that there would have been no habitable period (Hamano et al., 2013). Also, Venus’ present surface is geologically young (300–600 million years old; Strom et al., 1994). It has relatively few craters, and they are fairly evenly geographically distributed, which led to the hypothesis that the venusian lithosphere experienced one or more catastrophic overturn events, resurfacing everything at once (Strom et al., 1994). Thus, there is no geological evidence remaining of Venus’ more distant past.

While surface environments on Venus are evidently not habitable because of high temperature and pressures, there
has been speculation about life high in the atmosphere, where these parameters are moderate (Morowitz and Sagan, 1967). Life in certain layers of the atmosphere is generically possible for all planets with thick atmospheres. The major challenge for organisms would be to remain buoyant throughout their life cycles. The original suggestions were presented in the context of the misconception that Venus’ clouds were primarily made of water; without this, the challenges are greater, but speculation remains (Schulze-Makuch et al., 2004).

With its water lost, Venus is not considered to be habitable today. But the question arises: Was early Venus habitable? Did plate tectonics occur on Venus? When did the runaway greenhouse occur? If this occurred late in the history of Venus, the planet was probably habitable early in its history.

Recent reviews on Venus include discussion of its atmosphere (Hunten, 2007; Fegley, 2009), geochemistry (Fegley, 2003), surface and interior (Smrekar and Stofan, 2007), and the runaway greenhouse (Goldblatt and Watson, 2012). Grinspoon’s (1997) popular science book on Venus is comprehensive and accessible to the nonspecialist.

6.1.2. What do we know about the habitability of Mars? Was it ever inhabited? The Noachian era (4.5–3.7 Ga) was a period on Mars when the planet may have been relatively warm and wet, or alternatively cold with melting ice (Fashtook and Head, 2015). Some degree of liquid water on the surface is suggested by several lines of evidence (Fig. 19), which include the following: (1) the presence of ancient, water-eroded structures; (2) weathered exhumed phyllosilicates (clay minerals, Carter et al., 2010); (3) the density of valley networks likely produced by precipitation (Hynek et al., 2010); (4) salt deposits partly connected to valleys, which suggests water-related sedimentation (Osterloo et al., 2008); and (5) aqueous minerals such as jarosite and hematite observed by the Opportunity rover (Squyres and Knoll, 2005).

Noachian-aged crust also has magnetized stripes, which are thought to be remnants of an ancient magnetic field (Connerney et al., 1999). This magnetic field would have served as a shield against the solar and galactic cosmic rays, thus increasing the habitability of the surface of Mars.

During the Hesperian era (3.7–3.0 Ga), global climate changes resulted in a generally colder and dryer environment. Episodic water outbursts from subsurface aquifers may have occurred, however, and produced temporarily wetter conditions locally and/or regionally. Orbiting spacecraft observations of thick sulfate sediments in Valles Marineris and other locations (Gendrin et al., 2005) indicate the presence of relatively cold (in some cases below 0°C) and acidic waters in the Hesperian or Amazonian (Okubo et al., 2008). Beyond this simplified picture, a diversity of conditions is suggested by the occurrence of various minerals in this era, for example, diagenetic crystalline hematite and opaline, which likely formed through groundwater altering sediments (Christensen et al., 2000; Milliken et al., 2008).

Water has been scarce during the Amazonian period on Mars (3.0 Ga to the present). Signs of water during the Amazonian include channels on the flanks of some volcanoes and signs of ephemeral lakes at lower latitudes (Fassett and Head, 2008; Kereszturi, 2010). Gullies and some flowlike features (commonly referred to as “recurring slope lineae”) are thought to originate at midlatitudes from the melting of either an ice-rich layer that covers the surface or accumulated snow packs (Fig. 19). These are the best candidates for the most recent liquid water on the surface. Besides these proposed macroscopic signatures, thin water layers could exist between mineral grains in contact with ice at temperatures as low as –93°C (180 K, Möhlmann, 2004; Fig. 19). Such layers would be only a few nanometers thick and probably not useful for biological processes, although there is ongoing discussion about this topic (Möhlmann, 2009). Thicker films of adsorbed
water or larger occurrences of brines (salty water) could exist below 0°C (discussed more below).

The present surface conditions are harsh on the Red Planet; experiments in Mars simulation chambers (de Vera et al., 2010) suggest that survival in a dormant state is possible for certain microbes under radiation shielding. Achieving active metabolism for these microbes, however, requires a combination of accessible water and elevated temperature, conditions that might be very rare on Mars today. Water with dissolved salts, generically called brines, remains liquid at lower temperatures and evaporates at a lower rate than pure liquid water, so some have speculated that this could increase the habitable potential of environments on Mars. The change in size and albedo of droplet-shaped forms seen on the strut of the Phoenix lander (Rennó et al., 2009) presents evidence that brines may be associated with subsurface ice on Mars. Possible wet environments on Mars today include hygroscopic minerals (that take in and retain H2O molecules from the atmosphere; Fig. 19), subsurface ice overlain and protected by a porous soil layer, and seasonal changes possibly attributed to a liquid, flowlike process (Kereszturi et al., 2009; McEwen et al., 2011) as indicated by dark spots that appear on some dunes in springtime (Szathmáry et al., 2007). Subsurface caves serve as a specific type of habitat on Earth (Northup et al., 2011), and on Mars caves and old lava tubes could provide radiation shielding (Léveillé and Datta, 2010) and potentially trap ice (Williams et al., 2010). Although the temperatures would be low in these caves, they could still be important astrobiology targets for future missions.

The present-day martian surface does not appear to be habitable. The thin atmosphere does not protect the surface from the high doses of incoming solar radiation. Except for the possibility of thin and/or salty ephemeral liquid layers (Kereszturi and Rivera-Valentin, 2012; Martin-Torres et al., 2015), the very dry atmosphere prohibits the formation of bulk liquid. Challenges for any martian organism on the surface today would include the need for H2O uptake (possibly from salty water) and the fact that low temperatures inhibit metabolism, which is needed for ongoing repair of radiation damage and protection against the strong oxidizing agents such as perchlorate, which was detected by the Phoenix mission (Hecht et al., 2009). Subsurface environments, however, are more protected from other harsh surface conditions. A millimeter-thick rock layer provides shielding against solar UV radiation, while a meter is needed to block higher-energy particles. If subsurface environments incorporate a redox gradient with oxidized substances such as perchlorate from the surface, and more reduced substances from below, then that redox gradient could provide energy sources potentially suitable to sustain life (Link et al., 2005; see also Chapters 4 and 5).

Conditions may have been somewhat more habitable in Mars’ recent past. Geothermal centers might have been warm enough to melt ice (Fig. 19), and age estimations of some volcanoes show activity 30–10 million years ago, which would have allowed for the possibility of dormant, though not exhausted, volcanic activity. Observations of methane in the martian atmosphere suggest that some process is currently releasing methane, given that this is a gas that is relatively easily photodissociated (Formisano et al., 2004; Krasnopolsky, 2006; Mumma et al., 2009; Fonti and Marzo, 2010; Webster et al., 2015). The variability reported in both the remote (Mumma et al., 2009) and in situ (Webster et al., 2015) measurements of methane is difficult to reconcile with models of the chemistry network of the martian atmosphere. This indicates that either the measurements are wrong or our understanding of the chemistry of the martian atmosphere is incomplete. Observed atmospheric methane could be consistent with a variety of mechanisms that include subsurface biological methane production (Krasnopolsky et al., 2004), volcanic release, release from ancient sediments and clathrates (ice structures that capture gases and release the gas when destabilized) (Onstott et al., 2006), or serpentinization (a geological methane source observed on Earth; Sherwood Lollar et al., 1993). However, it is difficult to discriminate between these potential sources and sinks as long as there is a disconnect between models and observations. Future observations and missions will gather data that bear on this question (see Section 6.1.3), and future models will attempt to simulate these data by incorporating the processes above. A comparison of the geological eras of Venus, Earth, and Mars is shown in Fig. 20.

6.1.3. Inner Solar System missions. All planetary missions contribute to our understanding of the Solar System and, ultimately, how it formed. Here, we focus on a selected set of missions that have a direct astrobiological motivation. Future missions are those in development or recommended by the Planetary Science Decadal Survey for 2013–2022 (Visions and Voyages for Planetary Science in the Decade 2013–2022, by the Committee on the Planetary Science Decadal Survey, 2011) or the European Space Agency’s (ESA’s) Cosmic Vision 2015–2025 plan. The Planetary Science Decadal Survey gathers input from the scientific community and prioritizes the Solar System locations that should be visited with spacecraft in the coming decade. NASA has three general classes of missions, in order of scale and price: the largest are the Flagship missions, intermediate are New Frontiers missions, and smallest are Discovery missions. The decadal survey specifies locations where the first two mission classes can be sent, but the Discovery class is open to any proposals that fit within the budget constraints.

Some of the early planetary missions designed to specifically investigate habitability were the Viking missions to Mars in 1976 (see Section 7.1.3.7). Many missions have since visited Mars and investigated the atmosphere and surface with the intent to learn more about the planet’s evolution and potential habitability. The solar-powered Mars Exploration Rover (MER) named Opportunity is still roving the surface and learning new things about Mars (a notable feat given it landed in early 2004 with a nominal mission time of 3 months). Five missions are still active in orbit around Mars: Mars Express, Mars Odyssey, and Mars Reconnaissance Orbiter (see links to NASA and ESA sites below for details) and the Indian Mars Orbiter Mission, also called Mangalyaan (see the website of the Indian Space Research Organization for more information, www.isro.gov.in). The NASA Mars Atmosphere and Volatile Evolution (MAVEN) mission launched in November 2013 and arrived at Mars in late 2014. This mission is studying the upper atmosphere and its interactions with the solar wind to better understand atmospheric escape processes. This understanding is also key to understanding the evolution of the atmosphere and the possibility of past surface water on Mars [see the Science special issue of 6
November 2015 (Vol. 350, Issue 6261), which covers MAVEN’s early findings.

The Mars Science Laboratory (MSL) is the most recently landed rover. The rover, named Curiosity, carries multiple camera systems; several spectrometer packages (including a laser-induced breakdown spectrometer that can target samples from up to 7 m away); and instruments that can remotely or by way of collected samples characterize minerals, organics, and gases. This suite of instruments has been designed to give a more complete understanding of the past habitability of the martian surface and the sources and sinks for organic material on the modern surface. Organics (Freissinet et al., 2015), methane (Webster et al., 2015), and indigenous nitrogen (Stern et al., 2015) have been reported in the modern environment. Analyses of the geochemistry and sedimentology of martian rocks indicate an ancient lake bed that persisted for millions of years (Grotzinger et al., 2014), as well as flowing water (Williams et al., 2013). Curiosity has also constrained the long-term evolution of the martian atmosphere by measuring the D/H ratio (Mahaffy et al., 2015) of a rock that the rover also dated (Farley et al., 2014). Curiosity is still operating on the surface and at the time of this writing is analyzing beds of increasing age and varying composition in an attempt to build a historical context for its measurements.

A series of missions are also in development. ESA’s ExoMars/Trace Gas Orbiter (EMTGO), launched in 2016, will be followed in 2020 by the ExoMars rover, an ESA-led mission in collaboration with the Russian Federal Space Agency (Roscosmos). EMTGO will investigate the chemical composition of the martian atmosphere and search for signs of the disputed methane signature (see Section 6.1.2) as well as other atmospheric species (Gianfiglio, 2011). A Flagship class mission is also under development by NASA, designed to identify and cache samples on the martian surface for return to Earth via a subsequent mission. Sample return will involve several steps (the samples need to be collected, cached, and later returned to Earth). The InSight mission, a NASA Discovery mission scheduled to launch in 2018, comprises a lander equipped with geophysical instruments (including a seismometer and a heat flow probe) to investigate Mars’ interior. Taken together, this lineup of current and future missions will continue to develop a systems framework for understanding modern and ancient Mars, including information about its interior, surface, atmosphere, and space weather environment.

ESA’s Venus Express spacecraft was in orbit around Venus from 2006 to the beginning of 2015, and its main science goals centered on observations of the atmosphere. The atmosphere is tied to the surface and Venus’ past. Venus Express also took pictures of Earth as a way to study the characterization of a habitable planet from less than one pixel of information, which is an analogue to studies of extrasolar planets.

A few spacecraft have visited and investigated asteroids, which will aid in our understanding of planet formation and the origin of organics found in meteorites. The Dawn mission spent time investigating Vesta before arriving at its second target, Ceres, which is the largest object in the asteroid belt. The Hayabusa spacecraft, a JAXA mission, visited the asteroid Itokawa. This was the first spacecraft to land and take off from the surface of an asteroid and was able to return samples to Earth after a series of difficult engineering setbacks. The ESA Rosetta mission investigated the comet 67P/Churyumov–Gerasimenko and, along with its lander Philae (which made the first-ever landing on a comet), returned new information on long-period comets. This included measurement of the D/H ratio of the comet, the result of which is inconsistent with the hypothesis that the water in Earth’s oceans was delivered entirely by comets; see the Science special issues of 23 January 2015 (Vol. 347, Issue 6220) and 31 July 2015 (Vol. 349, Issue 6247) for details of Rosetta’s
early findings. Rosetta has also detected the amino acid glycine, and phosphorus, both essential for life on Earth, in the comet (Altwegg et al., 2016).

A second mission to return a sample from an asteroid (OSIRIS-REx, an acronym for Origins, Spectral Interpretation, Resource Identification, Security–Regolith Explorer) is currently under development at the New Frontiers level (NASA’s medium-sized missions). OSIRIS-REx is planned for launch in the fall of 2016. This craft will visit the ~500 m diameter asteroid Bennu (which is on a trajectory that has a small possibility of resulting in an impact with Earth in the late 22nd century) and bring back samples in a manner that will preserve organics. These samples could reveal information about Solar System formation and the source of organics available for early life on Earth (Steigerwald and Keesey, 2010). The mission should also enable greater understanding of Bennu’s trajectory to help determine whether or not it will one day hit Earth.

Many nations have space programs, and more information about planetary missions can be found at the various agencies’ websites (listed alphabetically):

- CNSA: http://www.cnsa.gov.cn/n615709/cindex.html
- ESA: http://sci.esa.int/science-e/www/area/index.cfm?fareaid=7
- ISRO: www.isro.gov.in
- JAXA: http://www.jaxa.jp/projects/sas/planetary/index_e.html
- NASA: http://science.nasa.gov/planetary-science/missions

FIG. 21. (Top) Predicted interior structure of Europa, showing it is composed of (from the inside out) a metallic core, a rocky mantle, a global subsurface ocean and outer ice shell. (Bottom) Close-up of the outer ~100–150 km of Europa illustrating the global ocean and overlying ice shell (kilometers to tens of kilometers thick). This ocean is potentially heated by hydrothermal activity and may be an abode for life. Europa is the fourth-largest satellite of Jupiter, with a radius of ~1565 km, and it has a surface area similar to that of the continent of Africa. (Credit: NASA/JPL)
6.2. What do we know about the habitability of the outer Solar System?

6.2.1. Is Europa habitable? The Galileo mission detected an induced magnetic field in Europa (a moon of Jupiter), providing compelling evidence for an ocean of liquid water beneath the surface (Kivelson et al., 2000). Jupiter’s magnetic field continually sweeps through Europa because Jupiter rotates faster than Europa orbits around it; this causes the induced field. This observation requires the presence of a conductive layer within 200 km of the surface, which is interpreted to be a salty global ocean below the outer icy shell. In addition, observations of surface fractures from Galileo imaging data reveal patterns consistent with an ice shell fully decoupled from the rocky core, and a broad range of geologic evidence for the presence of an internal ocean includes impact morphologies, cryovolcanic features, pull-apart bands, and topography (Pappalardo et al., 1999).

While the top surface of Europa is not habitable, it is still relevant to discuss here for several reasons: (1) surface materials may reflect the chemistry of a subsurface ocean and help us understand the conditions available to life there, and (2) materials brought to the surface from the ocean below may be the only way we can detect signs of life (e.g., organic matter or other biosignatures characterized spectroscopically on a future mission; Kargel et al., 2000; Dalton et al., 2003). Non-ice substances were observed on the surface by the NIMS imaging spectrometer on Galileo. Close spectral matches for these materials include hydrated sulfate minerals (McCord et al., 1998) and sulfuric acid hydrate (Carlson et al., 1999). Radiation from high-energy particles caught in Jupiter’s magnetic field causes oxidation of surface materials as the energy from radiation knocks electrons out of atoms. Radiolytically produced oxidants at the surface include O2, SO2, H2O2, and CO2 (Hand et al., 2005). These may be subsequently cycled into the ice shell and ocean on timescales of the order of the age of the surface, ~10^6 years, which is quite young on geological timescales (Hand et al., 2007), so these compounds could potentially oxidize more reduced chemical species in the ocean below.

Habitability within the ice layer may be possible if geophysical processes can drive temperatures high enough to allow for the presence of life. Geologic features on the surface of Europa (e.g., chaos regions, domes, and pits) point to the presence of convective activity within the ice layer (e.g., Pappalardo and Barr, 2004). This convective activity suggests thermal circulation and/or tidal heating, two processes that could provide temperature zones near or at the melting point of the ice within the shell. Europa is in an orbital resonance with its neighboring moons, Io and Ganymede. For every time Ganymede, the outermost large moon, completes one orbit around Jupiter, Europa completes two orbits, and Io, the innermost large moon, completes four. Gravitational interactions with Ganymede and Io keep Europa in a noncircular orbit, which means Europa’s distance from Jupiter changes over the course of its orbit, Jupiter’s gravitational attraction changes, and Europa’s tidal bulge raised by Jupiter goes up and down. This flexing causes the tidal heating that allows Europa to have active geology. Models predict that, when ice is locally heated at the base of the shell (by plumes of heat rising through the ocean from the rocky interior), it will rise due to thermal buoyancy, as warmer ice is less dense than colder ice. The ice will be further tidally heated as it rises, as warm ice is also more deformable than the surrounding cold ice. This interaction between convection (physical mixing caused by heating from below) and tidal heating enhances temperatures within upwelling plumes, which can even reach the melting temperature of ice, near 0°C (Barr and Showman, 2009, and references therein; Sotin et al., 2009, and references therein; Han and Showman, 2010; Schmidt et al., 2011). Metabolic activity in psychrophilic (“cold-loving”) organisms that live in permafrost and sea ice on Earth has been demonstrated down to ~−30°C (Rivkina et al., 2000; Joyce et al., 2004). Additionally, ice that contains impurities can retain liquid brines along ice grain boundaries; bacteria have been imaged at such interfaces on Earth (Deming, 2002).

Because it is a liquid water environment, the ocean is the most likely habitat for life in the euroman setting. Modeled salinity values, based on empirical calculations that compare observations from the Galileo spacecraft magnetometer-derived oceanic conductivities and radio Doppler data-derived interior models to experimental laboratory evaluation of conductivity versus concentration data for both magnesium sulfate solutions and terrestrial seawater, are well within known limits for Earth-based life (Hand and Chyba, 2007). Models of the interior structure of Europa (based on its size, overall density, and moment of inertia, along with other Galileo spacecraft measurements) predict that the ocean is in direct contact with the rocky mantle (Schubert et al., 2004; see Fig. 21) such that it facilitates water-rock reactions and perhaps interaction with hydrothermal vent systems. Such systems could introduce reductants (electron donors) to the ocean (McCollom, 1999) and drive prebiotic synthesis of simple organics (Russell and Hall, 1997). Organics could also be derived from cometary infall (Pierazzo and Chyba, 2002) or radiolytic activity at the surface (Moore et al., 2007). The seafloor-generated reductants could be coupled to surface-derived oxidants, setting up geochemical gradients that life could harness to gain energy (Zolotov and Shock, 2003, 2004; see Section 4.2.1). The overall H2O layer on Europa (outer ice shell plus ocean) is estimated to be 100–150 km (Schubert et al., 2004). Although the thickness of the ice shell is debated, it is estimated to be on the order of 10 km, while the briny ocean (liquid or at least slushy) part of the H2O layer is estimated to be on the order of 100 km. While pressures at the base of a euroman ocean exceed those of the deepest ocean trenches on Earth, we have not yet found an environment on Earth where the pressure is too great for life to survive.

Piezophilic (“pressure-loving”) organisms have been isolated from sediments sampled from the Mariana Trench and many other high-pressure marine locations on Earth (Nogi et al., 1998; Fang et al., 2010).

6.2.2. Is Titan habitable? Titan, Saturn’s largest moon, is a world rich in organic molecules and considered to be potentially habitable. Reactions with solar UV light and charged particles occur in its dense, nitrogen-methane atmosphere (~98% N2, 2% CH4) and produce a wide range of nitrogen-bearing organic molecules (Cuí et al., 2009; Hörst et al., 2012). These molecules subsequently rain down onto the satellite and coat its surface. The processes that occur on Titan may be similar to those that occurred on early Earth due to atmospheric chemistry (Trainer et al., 2006), which makes Titan an
interesting natural laboratory for the study of prebiotic chemistry. However, the lack of oxygen-containing molecules in Titan’s atmosphere limits the planet’s usefulness in this role, given that oxygen atoms are present in most known biological molecules. Small amounts of CO (~ 50 ppm) are present, so it is possible that photochemical processes may lead to the production of a limited amount of prebiotic molecules in Titan’s atmosphere (Hörst et al., 2012).

However, there is another, more abundant source of oxygen atoms on Titan: frozen and liquid water. Theoretical models of Titan’s formation and evolution (e.g., Tobie et al., 2006) predict that its interior consists of a water ice or methane clathrate layer several tens of kilometers thick that overlies a subsurface water ocean several hundred kilometers deep (Fig. 22). The response of Titan to tidal forces from Saturn is also consistent with the presence of a global ocean at depth (Iess et al., 2012). Although liquid water may be abundant inside Titan, it is frozen near the surface. For the organic molecules on Titan’s surface to interact and form biologically relevant compounds, liquid water, not ice, must be present. Titan’s surface temperature of 94 K (-180°C) means that Titan is, on average, too cold for liquid water to persist at its surface, but melting caused by impacts or icy volcanism may lead to its episodic availability. Impact melt pools on Titan would likely remain liquid for ~10² to 10⁴ years before freezing (O’Brien et al., 2005). When these two ingredients—complex atmospheric organics and liquid water—are combined, they have been shown to produce prebiotic compounds such as amino acids (Khare et al., 1986; Neish et al., 2010) on a timescale of days to months (Neish et al., 2008).

But has the chemistry on Titan moved beyond the creation of simple biomolecules to actual biology? There are two possible places where life might have originated on Titan—in a subsurface water ocean or on its surface. During Titan’s early history, its subsurface ocean was exposed to space, and impactors could have provided the nutrients needed for life, while radiogenic heat flow from Titan’s interior could have provided an energy source for life to utilize. This ocean has long since frozen over, but Fortes (2000) found that there are terrestrial organisms that can withstand the pressure (1–4.5 kbar), temperature (<260 K), alkalinity (pH <11.4), and viscosity (<6.4 centipoise) predicted for Titan’s subsurface ocean.

There are a variety of proposed habitats for life on the surface of present-day Titan as well. There may be local cryovolcanic hot spots where Earth-like, water-based life could currently survive (Schulze-Makuch and Grinspoon, 2005). Another theory is that liquid hydrocarbons could act as a food source. McKay and Smith (2005) calculated the energy available for organisms on Titan utilizing such a scheme. They predicted that titanian organisms could “make a living” by consuming hydrogen (H₂), acetylene (C₂H₂), and ethane (C₂H₆), generating methane (CH₄) as a waste product. Another potential habitat on Titan is its

FIG. 22. (Left) Titan, as seen by the Cassini spacecraft in UV and IR wavelengths, which revealed the surface through Titan’s dense atmospheric haze. (Right) The surface of Titan as viewed by the Huygens descent probe. The rounded appearance of the pebbles is consistent with long-distance transport along a riverbed (Tomasko et al., 2005). The surface area of Titan is approximately twice that of the continent of Asia. (Credits: NASA/JPL/ESA/Space Science Institute/University of Arizona)
extensive polar lakes and seas, which are almost certainly composed of liquid methane and ethane (e.g., Brown et al., 2008). These environments provide a natural test bed for the idea that life requires water as a solvent. The detection of complex polymers in these reservoirs—possibly with a “lake lander” such as the Titan Mare Explorer (Stofan et al., 2013)—would provide evidence for chemical self-organization in non-aqueous environments (Lunine, 2014). However, the very low temperatures on Titan make it difficult for any life on the planet to survive, from an energetic standpoint (see Section 4.2 for more on energy and life).

6.2.3. Is Enceladus habitable? Enceladus is a small moon of Saturn. Based on its size alone, Enceladus would be expected to have a cold interior, to exhibit minimal geological activity, and would probably be of little interest to astrobiologists. However, Enceladus does not conform to these expectations. Rather, we have learned that Enceladus is a dynamic world, featuring a south polar region that has been ravaged by tectonic forces and exhibits a towering plume of gas and misty ice (Fig. 23). In terms of geophysics, Enceladus is undoubtedly alive. While it is not yet known whether Enceladus is also habitable, the interior of Enceladus may provide three key requirements for life: liquid water, nutrients, and energy gradients.

The inferred source of Enceladus’ plumes is a liquid water ocean (Porco et al., 2006). Measurements of Enceladus’ gravity field and physical liberation (a slight wobble in its orbit) imply a global subsurface ocean (Iess et al., 2014; Thomas et al., 2016). The presence of an internal ocean allows enhanced tidal heating that is needed to explain high heat fluxes from the south polar region (Tobie et al., 2008). An ocean source of the plume is supported by the detection of salts, such as NaCl, in ice grains in the plume (Postberg et al., 2011). This argues for water-rock reactions to leach soluble salts from Enceladus’ rocky core. Specifically, geochemical modeling suggests that serpentinization (hydration of ultramafic rock) may be occurring, which would produce alkaline water (Glein et al., 2015), and this process may be occurring at elevated temperatures (>90°C) at least locally and producing silica nanoparticles (Hsu et al., 2015).

It is likely that the chemical elements used by life on Earth (the CHNOPS elements) are present in Enceladus’ ocean or core. Indeed, gaseous H2O, CO2, CH4, NH3, and perhaps H2S were detected during multiple flybys of Enceladus when the Cassini spacecraft flew through the plume (Waite et al., 2009). More massive organic compounds also appeared to be present, but their abundances were low, making identifications difficult. Potential sources of phosphorus and sulfur include minerals in the core, such as apatite and iron sulfides.

There could be enough chemical disequilibrium inside Enceladus to support microbial ecosystems similar to some on Earth. One possibility is that serpentinization generates H2, the presence of which would allow organisms to biologically produce CH4 (or acetate) from CO2 (McKay et al., 2008). CH4 was observed in the plume, although its origin is unknown because of the lack of isotopic data. A key issue with regard to the geochemical support of contemporary life on Enceladus is whether serpentinization and H2 production are ongoing. A major objective of the final Cassini flyby is to search for H2 in the plume that could come from hydrothermal vents (Hsu et al., 2015). An alternative source of metabolic
energy could be provided by the dissociation of liquid water from radioactive decay in Enceladus’ core or irradiation of water ice on its surface (McKay et al., 2008). These processes are robust with respect to time and should occur today. However, the oxidants (e.g., sulfate, H$_2$O$_2$, O$_2$) that would be produced have not been detected in the plume or on the surface. This could be explained by slow rates of production (low radiation doses) and/or rapid removal processes that prevent their buildup (including biological consumption, such as sulfate reduction). Unraveling the redox chemistry of Enceladus is critical to improving our ability to assess the habitability of this ocean world.

6.2.4. Are other icy moons habitable? In the search for other habitable environments in the asteroid belt and outer Solar System, there is limited data with which to work. However, one key criterion to be considered, both with observations and theory, is the potential presence of liquid (water or otherwise) on these bodies. In the outer Solar System, this usually equates to places with signs of past or present geological activity.

Few outer Solar System moons have been proven to be currently active. Besides the plumes on Enceladus (see Section 6.2.3), Neptune’s moon Triton also has current activity. The Voyager 2 spacecraft observed active N$_2$ plumes (Smith et al., 1989; Buratti et al., 1999), which contribute to Triton’s tenuous nitrogen atmosphere. This nitrogen condenses on the surface, along with CO, CO$_2$, and CH$_4$ ices that have been identified from ground-based spectroscopy (Grundy et al., 2010, and references therein). Triton has relatively few craters, indicating a geologically young surface. Many areas of Triton have morphological features similar to those caused by icy volcanism: pits, vents, caldera-like complexes, and smooth surfaces (Fig. 24). In models, this icy volcanism, or cryovolcanism, is often facilitated by NH$_3$ or other compounds that lower the melting point of ice (McKinnon and Kirk, 2007). However, NH$_3$ has not yet been observed on Triton’s surface.

The same internal heat that allows Triton to have geologically recent surface activity may also allow for a subsurface ocean. Models predict that a liquid layer might be sandwiched between an outer ice shell and an inner layer of high-pressure phases of ice (McKinnon et al., 1995). If an inner icy layer underlays the ocean, it would mean that the ocean would not be in contact with the inner rocky mantle (unlike the ocean on Europa, which is in contact with the mantle). Without a continuous source of reduced or oxidized compounds to create a redox gradient (such as might be

![FIG. 24. Neptune’s moon Triton exhibits active nitrogen plumes and exotic surface geology, such as icy volcanic features, and the hummocky cantaloupe terrain seen at upper left. The same internal energy that allowed for recent geological activity also creates the possibility for a subsurface ocean on Triton. The radius of Triton is 1350 km, giving it a surface area equivalent to that of North America. (Credits: NASA/JPL/USGS and Smith et al., 1989)](image-url)
produced by hydrothermal vents on Europa), it is less obvious what the source of energy for potential tritontarian organisms could be. The ocean would also occur at a much deeper level (hundreds of kilometers under the surface) than the ocean on Europa. Due to this, communication between the ocean and the surface is likely less efficient. Thus, this type of subsurface ocean is considered more theoretically challenging to describe as habitable, at least for the requirements of Earth-like life. Jupiter’s moons Ganymede and Callisto and Saturn’s moon Titan (see Section 6.2.2) likely also have internal oceans, but similar to Triton, they are not necessarily predicted to be in contact with a rocky mantle. Modeling of Ganymede’s interior structure suggests it may have layers of different phases of ice and water, which could allow for some water-rock interactions, but it is unclear whether any products created would be mobile throughout the shell (Vance et al., 2014).

6.2.5. What is the water inventory of small bodies in the Solar System? In addition to the icy satellites, other small bodies in the Solar System may be capable of presently harboring liquid water and can provide clues to the Solar System’s formation and history. Modeling of large trans-neptunian objects, bodies that lie beyond the orbit of Neptune, shows that heating from the long-lived radiogenic nuclides may provide enough warmth to sustain liquid water if there are antifreezing species such as ammonia present (Hussmann et al., 2006; Desch et al., 2009). Given the overwhelming population of the Kuiper Belt (10^6 objects greater than a few tens of kilometers; Morbidelli and Levison, 2007) and Oort Cloud (~4×10^11 objects with a radius larger than 1 km; Levison et al., 2010), trans-neptunian objects may provide a significant fraction of the Solar System’s liquid water reservoir, perhaps collectively even as much as on Earth (Desch et al., 2009). While these bodies would be a challenging place for life to develop and be supported, water-rock interactions at the boundary of a warm silicate core (if the body is large enough) and a liquid water ocean might provide a relatively hospitable environment. Indeed, young surfaces on Pluto were observed by New Horizons in 2015, illustrating that even in the far reaches of the outer Solar System, small differentiated objects can remain geologically—or cryovolcanically—active over long periods.

Comets are another class of small bodies that are reservoirs of water in the Solar System. They are composed largely of dust and water ice that may have remained a relatively pristine example of our early solar system. As such, they are cryogenics to the earliest stages of solar system evolution. Comets are thought to have impacted early Earth, and it has been proposed that they may have delivered a significant volume of water (in the form of ice) and other volatiles to the planet (e.g., Laufer et al., 1999; Delsemme, 2000; Levison et al., 2001). Comets are also rich in organic material that may have contributed to the rise of life on Earth (e.g., Delsemme, 2000; Altwegg et al., 2016; see Sections 3.1.1 and 6.1.3).

Most asteroids are small, predominantly rocky bodies, but it has long been known that some asteroids likely contained a greater volume of water early in Solar System history. We know this from the study of certain meteorites that contain minerals altered by liquid water. The parent bodies that broke apart—forming debris, some of which later fell on Earth as meteorites—are believed to have undergone intense hydrothermal activity. This hydrothermal activity would have been a sink for the water. Depending on the extent of hydrothermal activity, water on an asteroid may be consumed by water-rock reactions, or some of it may remain in a thin outer layer enriched in water, hydrates, and organics (Castillo-Rogez and Schmidt, 2010).

At nearly 1000 km across, Ceres is the largest object in the asteroid belt between Mars and Jupiter. It is 25% the size of our Moon and twice the size of Enceladus and likely differentiated into a rocky core and icy shell (McCord and Sotin, 2005; Thomas et al., 2005). Ceres, which was visited by the Dawn spacecraft in 2015, has a composition that may include as much as 3 times the volume of water in Earth’s oceans, a portion of which may remain melted in a global subsurface ocean even today (Castillo-Rogez and McCord, 2010). Understanding water and organics in asteroids has important implications for the evolution of life in the early Solar System and delivery of these materials to Earth (see Sections 3.2.1 and 3.2.2).

6.2.6. Outer Solar System missions. Here, we briefly discuss past and potential future astrobiologically relevant missions to bodies beyond the asteroid belt. Future missions included in this discussion are those recommended by either the Planetary Science Decadal Survey for 2013–2022 (Committee on the Planetary Science Decadal Survey, 2011) or ESA’s Cosmic Vision 2015–2025 plan. See Section 6.1.3 for a description of NASA mission classes.

The Galileo mission orbited Jupiter from 1995 to 2003. It carried a large suite of instruments to study both Jupiter and its moons, including Europa (see Section 6.2.1), and released a descent probe into Jupiter’s atmosphere. The Cassini-Huygens Mission (joint between NASA/ESA/ASI—the Italian Space Agency) has been in orbit in the saturnian system since 2004 and is planned to continue science operations until 2017. Astrobiology targets in the saturnian system that Cassini is studying include Enceladus and its south polar plume (see Section 6.2.3) as well as potentially habitable environments on the surface of Titan (see Section 6.2.2). Cassini also released the Huygens probe onto the titanian surface, the only probe to have landed on the surface of an outer Solar System moon (Fig. 22).

NASA’s Juno mission arrived at Jupiter in July 2016 and will study Jupiter’s atmosphere and deep structure to further our understanding of Jupiter’s origin and evolution. Europa and Enceladus are both recommended to be visited by Flagship class missions in the Planetary Science Decadal Survey if sufficient funds become available in the next decade (Committee on the Planetary Science Decadal Survey, 2011). The funding to start a mission to Europa was in NASA’s 2015 budget; this mission, the Europa Multiple Flyby Mission, would go into orbit around Jupiter so that it could make multiple flybys of Europa. Both Europa and Enceladus are worlds that have subsurface ocean environments, likely in contact with silicates, and spacecraft sent to orbit these moons could provide more insights into their potential habitability. Orbiting spacecraft would be able to take high-resolution images of the surface geology and gather compositional data as well as constrain their internal structures. For Europa, an orbiter could also scout out future
Another high-priority target for astrobiology in the outer Solar System is Saturn’s moon Titan, and several missions have been proposed to investigate potential hydrocarbon lake environments on the surface. In 2010, a mission called TiME, the Titan Mare Explorer (Stofan et al., 2010), was proposed (but not selected) in the 2012 Discovery class mission competition. This lake lander was proposed to investigate habitability by studying Titan’s active methane cycle and other, potentially prebiotic organic molecules, all while floating in one of Titan’s north polar seas.

In addition to missions targeting the icy moons, a comet sample return mission is recommended by the Decadal Survey as a New Frontiers class mission (Committee on the Planetary Science Decadal Survey, 2011). This would have similar objectives as an asteroid sample return mission (see Section 6.1.3) as far as preserving organic material for analyses that would help us understand Solar System formation and the early organics on Earth available for life to utilize. Comets also have the potential to reveal information about the history of volatiles and water delivery to Earth (see Section 6.2.4). A previous comet sample return mission, Stardust, flew by the comet Wild 2 and sent collected dust back to Earth in 2006. The samples have revealed interesting organics and information about mixing of material in the early Solar System (Brownlee et al., 2006). A mission concept where a spacecraft would rendezvous with and land on a comet multiple times and at different points in the comet’s orbit (called Comet Hopper or CHopper) has been proposed but not yet selected as a Discovery class mission.

More information about planetary missions can be found at the various agencies’ websites (listed in Section 6.1.3 above).

6.3. What do we know about the habitability of planets outside the Solar System?

6.3.1. What are the characteristics of known exoplanet systems, and how do we detect them? For most exoplanets, the characteristics that can be measured using current instrumentation include a lower limit on the mass of the planet, the radius of the planet, orbital eccentricity (how noncircular the orbit is), orbital period, and the average distance between the planet and the star. Other characteristics of the host star can be inferred: life stage or age, mass, radius, effective surface temperature, metallicity (described by the iron-to-hydrogen ratio, which indicates the relative abundance of heavier elements to lighter ones), and distance from the Sun. If both radial velocity and transit techniques are applied to the same target, additional information can be obtained on the mass (instead of minimum mass) and density of the planet, which in turn can lead to inferences on bulk planet composition. Over 1500 confirmed exoplanets have been catalogued at the time of publication of this primer, with several thousand additional unconfirmed planet candidates from the Kepler mission. Around 1000 of these planet candidates are in multiple-planet systems, with some systems having as many as seven known planets. (These data come from the Exoplanet Orbit Database, http://exoplanets.org. For updated numbers, see that website or a European database, the Extrasolar Planets Encyclopaedia, exoplanets.eu.)

Almost all discoveries and subsequent studies of extrasolar planets have been made through multiple indirect techniques. The five main indirect techniques are described here:

1. Radial velocity (also known as the Doppler method) is based on the fact that a star’s velocity toward and away from us varies cyclically due to the gravitational tug of an orbiting planet. This varying velocity is measured from shifting of the star’s spectral lines, and this observation allows estimation of the planet’s minimal mass and orbital period. The mass depends on the inclination of the system and is therefore only a minimal mass ($m \sin i$). If a transit of the planet is also observed, the inclination can be determined, and the observed minimal mass is indeed the planet’s mass.

2. Transit searches that look for dimming of the light from a star-planet system. The dimming occurs when an orbiting planet passes in front of its star from our viewpoint on Earth (called a transit) or behind the star (eclipse or occultation), and provides information on the orbital period and the planet’s radius. If multiple transits occur in the same system, their mutual effect on the timing of each others’ transits can be used to constrain the masses of planets in the system (this is called “transit-timing variation”).

3. Gravitational lensing, which relies on the lensing effect that concentrates light from background objects when a star system passes between it and Earth. The relative positioning of the star and planet changes the properties of the lens. Unlike most of the other techniques, gravitational lensing has an observational bias toward larger-period orbits, making it extremely complementary to the other detection techniques.

4. Pulsar timing, which is based on the fact that pulsars (aka neutron stars) emit regular pulses of radiation, whose timing will vary slightly due to the gravitational tug of a planet.

5. Astrometry, which involves measuring tiny changes in the position of a star as it moves in a very small elliptical orbit around the center of mass between it and an orbiting planet. This delivers information about the orbits and masses of the stars and planets in the system.

Direct techniques such as optical or IR imaging are becoming an increasingly important method for planet detection. Directly imaging the extremely dim light given off by an exoplanet is a technical challenge, especially for close-in planets (like those in the habitable zone; see Section 5.3.3 for a discussion of this term), because the light from the parent star is much brighter and needs to be blocked out in order to see the dim planet. But instrumentation and techniques are being developed to overcome this, and at the time of this writing over 60 planets have been imaged, primarily in the IR wavelengths. Advanced coronagraphs (instruments that block out the central star’s light) and the use of Earth or space-based interferometry (where the signal from many telescopes is combined to maximize observing capabilities) enable imaging of dim planets and precise measurements of
the positions and motions of stars. (See Section 6.3.3 for more discussion of current and future telescopes and missions searching for extrasolar planets.)

Originally, detection techniques were best at finding massive planets that orbit close to their host stars, creating a selection effect in the observations. As a result, many of these first exoplanets detected had a mass and planetary radius similar to or greater than that of Jupiter (Fig. 25) but were much closer to their host stars than Jupiter. Although not considered habitable, these “hot Jupiters” have challenged models for planet and star formation (see Sections 3.1.2 and 3.1.3) and in doing so have advanced our understanding of these processes.

As detection techniques have improved, astronomers have detected progressively smaller extrasolar planets, including “super-Earths”—planets with a mass up to 10 times the mass of Earth (or, equivalently, those whose radius is less than ~2 times larger than Earth’s). Once observational biases are taken into account, super-Earth planets are among the most common planet size based on detections thus far (Howard et al., 2012; Mullally et al., 2015). Many different factors would affect the habitability of these planets (see Section 6.3.2), but a few super-Earths have been found in what is thought to be the habitable zone of their host star, defined here as a region around the star where liquid water could exist on a planet’s surface (see Kopparapu et al., 2013; see Section 5.3).

Determining the density of a planet requires knowledge of both the planet’s size and mass. The planet’s size (diameter) is usually determined with the transit method, while the mass is usually measured by the radial velocity method. Other observations, such as measurements of changes to the timing of transits in multiplanet systems (“transit-timing variations”) can also constrain the mass. There are nine super-Earths with good determinations of both their radius and mass: GJ 1214 b (Charbonneau et al., 2009), 55 Cnc e (Gillon et al., 2012; Nelson et al., 2014), Kepler-36b (Carter et al., 2012), HD 97658 b (Dragomir et al., 2013), CoRoT-7b (Barros et al., 2014; Haywood et al., 2014), HIP 116454 b (Vanderburg et al., 2015), Kepler-78b (Howard et al., 2013; Pepe et al., 2013), and Kepler-10b and 10c (Dumusque et al., 2014).

The first potentially rocky planet (CoRoT-7b) was discovered in 2007 (Hatzes et al., 2011). This planet, though potentially rocky, is exposed to very high radiation from its extremely close host star and might have a lava ocean present on its surface, which vaporizes and forms its atmosphere (Schaefer and Fegley, 2009; Miguel et al., 2011). Since then, several terrestrial, rocky exoplanets have been detected. This has allowed the first attempts at creating mass-radius relationships for extrasolar planets (e.g., Wolfgang and Laughlin, 2011; Lopez and Fortney, 2013; Weiss and Marcy, 2014). The analyses of these relationships suggest a “break point” around 1.7 Earth radii (~5 Earth masses); above that size, increasingly larger planets have a composition that is increasingly dominated by H2/He gas.

Planets larger than Earth-sized and super-Earth planets, but smaller than the mass (or radius) of Neptune, are often

**FIG. 25.** Distribution of exoplanet semimajor axes (a measure of the distance of the planet from its star) and mass in terms of Jupiter masses. Large, gas giant planets are red circles, super-Earths to Earth-mass planets (1–10 Earth masses) are blue diamonds, and planets less massive than Earth are shown with green boxes. Additionally, the planets in the Solar System are marked for reference with open circles, and the yellow triangles represent the directly imaged HR 8799 system. Data are confirmed planets with mass estimates (http://exoplanet.eu, downloaded November 19, 2015).
referred to as “sub-Neptunes” or “mini-Neptunes.” While these seem to be among the most common types of planets in the Universe (or at least the portion we have explored), there are no examples of these worlds in our solar system. This prevalence outside our system, but absence inside it, makes them a primary target for future study. Multiple missions have been proposed or studied to observe them, and if NASA’s Astrophysics Division flies the Wide-Field Infrared Survey Telescope (WFIRST) mission with a coronagraph (Section 6.3.3), it should be able to analyze the chemistry of a few of these worlds. The James Webb Space Telescope (JWST) should also contribute to our understanding of the chemistries of these planets’ upper atmospheres if they transit nearby stars. Until further observations are made, our knowledge of these worlds comes from our limited observations and from various models that have been validated against the rocky planets, ice giants, and gas giants we have in our solar system (e.g., Line et al., 2011; Moses et al., 2013; Miguel and Kaltenegger, 2014).

NASA’s Kepler mission has discovered thousands of planet candidates (3786 according to exoplanets.org on May 30, 2016) using the photometric transit method, including planets similar in size to Earth (Fressin et al., 2012), planets potentially in the habitable zone, many multiplanet systems (Borucki et al., 2011; Batalha et al., 2012), and planets in binary star systems. Many of these are still considered “candidate” planets, as subsequent observations are required to rule out false positives and constrain the mass of the planet. The confirmation of a planet can come from measuring transit-timing variations, which tell us how the masses of different planets in the system affect the timing of each planet’s orbit, or from independent measurements from ground-based observatories (usually with the radial velocity technique). Either of these provides an independent means of verifying that the variations in the light curves are caused by an object in orbit around the star that has the mass of a planet (and not the mass of a star).

The Kepler mission team announced the discovery of the first rocky planet found by Kepler in January 2010, Kepler-10b, but this system is not in a habitable zone (Batalha et al., 2011). Kepler-22b, which was announced in December 2011, is in the habitable zone of a Sun-like star (Borucki et al., 2012). However, with a radius ~2.4 times that of Earth, it is unknown what the surface would be like, and it probably more resembles a mini-Neptune than a super-Earth (see Section 6.3.2 for a theoretical discussion of super-Earth surfaces). The Kepler-69 system is one of the most interesting, to date, from an astrobiological perspective. This system harbors two potentially habitable planets, Kepler-62e and Kepler-62f (Borucki et al., 2013). The smaller of these two planets (Kepler-62f) has a radius 40% larger than Earth’s and lies in the middle of the classic habitable zone. The Kepler 186 system is also of interest, because its fifth planet (Kepler-186f) is the smallest planet currently known to exist in another star’s habitable zone, with a radius only 11% larger than Earth’s. In 2015, a planet in the habitable zone around a Sun-like star was detected. Kepler-452b is 60% larger than Earth and slightly older and has been referred to as “Earth’s cousin” (NASA press release 15-156). Unfortunately, nearly all the potentially habitable Kepler planets are too far away for direct follow-up observations or radial velocity measurements; instead, future observations of these worlds will be limited to low-probability, directed searches for signals from extraterrestrial civilizations.

Such a search is already underway for some planets, including the aforementioned Kepler-186f. The existence of these planets is also relevant to the design of future missions that would attempt to observe potentially habitable worlds. The abundance of potentially habitable worlds ($\eta_{\text{Earth}}$)—where habitable is defined as planets smaller than gas giants, larger than atmosphere-free planets such as Mercury, and inside the habitable zone of their stars—will be a major factor in determining the yields from such missions. This, in turn, has a significant impact on the telescope aperture required to obtain a given sample size (Stark et al., 2014). As a result of the Kepler mission and a corresponding ground campaign, estimates for $\eta_{\text{Earth}}$ can now be based on observations instead of planet formation theory. However, a significant spread in these estimates exists due to effects of the host star type on planet abundances and disagreements on the size of the habitable zone, the range of planet sizes that can allow for global surface habitability, and the techniques used to account for biases in the observational data. A few exoplanets have been directly photographed (see http://apod.nasa.gov/apod/ap081114.html and http://apod.nasa.gov/apod/ap081117.html).

### 6.3.2. How can we characterize the habitability of exoplanets?

Here, we explore a few of the concepts that may affect the surface habitability of extrasolar planets. This discussion is by necessity restricted to the surface, because life in the subsurface of exoplanets is likely undetectable across interstellar distances. Both astrophysical and geo-physical perspectives bear on this topic, and there are many variables to consider when discussing the theoretical state of exoplanets. These parameters include both physical characteristics (e.g., mass, composition) and orbital characteristics (e.g., semimajor axis, eccentricity). Changing these parameters allows for a large variety of exoplanets, many of which are exotic when compared to our solar system. We will discuss some of these here. There is often a range of values for a given parameter that would theoretically be most ideal for habitability (e.g., the “Goldilocks zone” of semimajor axes described in Section 5.3.3, restricted to surface habitability here, and with respect to Earth-like life). To further complicate matters, these parameters are often interrelated and thus complex to model. Additionally, we have little data about the actual surfaces of these planets to constrain habitability models. As a result, many aspects of extrasolar habitability are still up for conjecture and debate.

Apart from liquid water at the surface, many other aspects will play into the surface habitability of exoplanets. Plate tectonics may place important constraints on surface habitability, as steady resurfacing of Earth has been shown to help stabilize the climate (Parnell, 2004) through the long-term carbon cycle (also see Section 4.2.2) and to recycle elements required for life. The mass of the planet might influence the potential for plate tectonics to develop, but recent model results related to this question are contradictory as to whether plate tectonics would be more or less likely on super-Earths (O’Neill et al., 2007; Valencia et al., 2007; Stein et al., 2011; Van Heck and Tackley, 2011; Noack and Breuer, 2014). The initiation of plate tectonics is an extremely complex topic and
not fully understood, but certainly many factors are at play (e.g., McKenzie, 1977; Mueller and Phillips, 1991; Stern, 2004; Korenaga, 2013). On Earth, the tectonic forces come from plate motions that are already occurring (e.g., Mueller and Phillips, 1991; Hall et al., 2003). However, on a stagnant-lid planet such as Venus, the first episode of plate tectonics must be initiated by other factors that allow or cause the lithosphere to fracture. Mantle convection has been proposed as a possible mechanism to initiate subduction, when one plate begins to sink into the mantle (Ogawa, 1990; Fowler and O’Brien, 2003; Solomatov, 2004), but a weak lithosphere is required (Toth and Gurnis, 1998; Wong and Solomatov, 2015). Water can be an effective agent to weaken the lithosphere so a lithosphere devoid of water may not be able to break, thus inhibiting the initiation of subduction (when one plate begins to sink into the mantle) and plate tectonics like that of Earth (e.g., Smrekar and Stefan, 2007; Lenardic et al., 2008, and references therein). An additional discussion point around the lack of plate tectonics on Venus involves the much warmer surface temperatures present on the planet, which could be similar for a hot exoplanet. Higher temperatures lead to lower mantle viscosities and thus lower convective stresses at depth in the lithosphere, which means plates may not fracture (Lenardic et al., 2008). Furthermore, at high temperatures, material that has been damaged can heal quickly by regrowth of the silicate grains. Thus, repair may occur on short timescales, and the lithosphere may not break into plates (Landuyt and Berкович, 2009).

Tidal heating has been proposed to affect habitability in that it may provide an additional source of internal heat to the planet (see Section 6.2.1 for an example and explanation of tidal heating). Most of Earth’s heat comes from the decay of radioactive elements such as uranium and thorium. For planets around small-mass stars, where the stellar insolation (light reaching the planet’s surface) is weaker, tidal heating could provide an additional source of heat and, thus, effectively widen the habitable zone (Barnes et al., 2009; Henning et al., 2009). Planets in multiplanet systems can have resonances similar to satellite systems in our own solar system that lead to tidal heating (Meyer and Wisdom, 2007). Additionally, many known exoplanets orbit close to their stars, and tidal heating should be intense on some of these worlds.

Too little internal heat can lead to an inactive geological interior, while too much heat can lead to extreme volcanism that can trigger a runaway greenhouse effect as occurred on Venus; neither situation is considered hospitable to life. Further, tidal forcings can also affect the climate of the planet by affecting the distribution of incoming starlight to the planet’s surface. A relatively well-known end-member example of this is when planets are close enough to their stars to become “tidally locked.” In the absence of torques from other planets, tidal locking would lead to a synchronous rotation where the same part of the planet is constantly irradiated by light, so there is a permanent “dayside” and a permanent “nightside” to the planet. Tidally locked planets have been suggested to be habitable at the boundary between night- and dayside, where temperatures might be in a habitable regime (i.e., allowing for liquid water at the surface). Joshi (2003) concluded that tidally locked planets should be habitable in general, although the habitability may depend on which part of the planet’s surface faces the star (Edison et al., 2012), and strong atmospheric circulation may be expected as a result of these intense temperature differences (between day- and nightsides), which could add additional challenges for surface habitability. However, if other planets exist in the system, planets that are tidally locked to their star may still receive sufficient torques from the other planets to prevent synchronous rotation and thus would avoid the development of permanent day- and nightsides.

Other orbital effects could also be important to a planet’s habitability (Spiegel et al., 2010). Planets with high obliquities (in other words, planets “turned on their sides”) may be habitable farther from their host stars, because this would cause the poles to be hit with large energy fluxes, which would suppress the runaway glaciation at the outer edge of the habitable zone (Spiegel et al., 2009; Armstrong et al., 2014). While eccentricity does not generally affect habitability (Williams and Pollard, 2002), it may become an issue when the eccentricity varies significantly or interacts with the varying obliquities (Dressing et al., 2010; Spiegel et al., 2010).

The existence of a moon around a planet or companion planets in a multiplanet system can also lead to orbital and tidal effects. The Earth’s Moon is thought to stabilize our obliquity, keeping Earth at the same tilt relative to the plane of the Solar System. The stability of the planet’s tilt stabilizes the equator-to-pole distribution of light on the planet and, therefore, also stabilizes the planet’s climate (Ward, 1982; Ward and Brownlee, 2000). For example, the obliquity of Mars has been much more chaotic over Solar System history, which is thought to have caused extreme changes in martian climate history (Touma and Wisdom, 1993).

Magnetic fields can also play a large role in surface habitability by shielding life at the surface from radiation and helping a planet retain an atmosphere. To maintain a magnetic field that is generated by the iron core of a terrestrial planet, a high heat flux at the core-mantle boundary is necessary, as is some threshold rotation speed (Schubert and Soderlund, 2011; Stevenson, 2003; and references therein). Depending on the planet, such a high heat flux might only be possible if the mantle of the planet is cooled efficiently by a mechanism like plate tectonics. For exoplanets much more massive than Earth, high interior pressures may hinder the development of a magnetic field (Elkins-Tanton and Seager, 2008; Barnes et al., 2010; Stamenskovic et al., 2012).

High pressure in the mantle of a large rocky planet can also prevent active volcanism at the surface by making melted rock negatively buoyant (Ottani et al., 1995). Such a pressure transition occurs on Earth at about 350 km in depth, allowing the rise of melt through much of the upper mantle, but only at ~100 km depth for a 10-Earth-mass planet (its much higher gravity leads to higher pressures at shallower depths). For large super-Earths, partial melt produced in the mantle may mostly stay below the lithosphere, and volcanic gases may contribute less to the composition of the atmosphere than volcanoes have throughout Earth’s history.

Exotic internal planetary structures and compositions (compared to the Solar System) can arise from various-sized planets that form in different parts of the stellar nebula or can arise from planets forming around stars with different overall abundances of elements than our sun. Earth-sized or super-Earth planets forming far enough from their stars for
ices to condense (see Section 3.1.3) may consist of large amounts of volatiles, leading to thick ice layers on top of these planets if they differentiate (Léger et al., 2004). The geological activity of such planets would strongly differ from a planet like Earth, because the ice layer would cut off the surface from the rocky mantle. They would likely not be habitable at their surface, because surface temperatures would be too low to allow for liquid water at this distance from the star. Possible subsurface oceans may occur, but subsurface life is more difficult to detect across interstellar distances, and no method for doing this has been developed. If these planets migrate closer to their stars (a common process in Solar System orbital dynamics), some ice may melt, forming an "ocean world" with a global ocean hundreds of kilometers deep (Léger et al., 2004; Sasselov and Valencia, 2010).

New data are revealing interesting dynamical situations that can further our understanding of planetary migration. The Kepler-20 system has two small Earth-sized planets tucked in between three large "ice giant"-type planets (large planets with significant amounts of water that are more similar to Uranus or Neptune than Jupiter), all closer to their star than Mercury is to the Sun (Fressin et al., 2012; Gautier et al., 2012). This configuration would not be predicted with current migration simulations (the smaller planets would likely be ejected). There is also a trend, in the planetary systems we have observed so far, for multiplanet systems to have more circular orbits than single-planet systems (Kane et al., 2012). Although there is the potential for observational bias against highly eccentric orbits, the long baseline (in terms of how long observations have been made using this method) of radial velocity measurements should be able to overcome this bias.

6.3.3. What are the current, future, and proposed missions to study extrasolar planets? Current space-based missions include the ESA Global Astrometric Interferometer for Astrophysics (GAIA) mission and the NASA Kepler mission. While GAIA uses astrometry to map the positions of stars and can thus detect giant exoplanets, Kepler conducted transit searches and contributed (together with the past ESA CoRoT mission) to estimating the occurrence rate of Earth-like planets around stars in the solar neighborhood and provided information on exoplanets across a wider set of parameters. Data from the original Kepler mission are still being analyzed, and new planets and planet candidates could be discovered based on past observations. There is also an extended "K2" mission that is still searching for planets with the Kepler telescope, but the current orbit and targeting strategy (which is limited by failed reaction wheels) will be mostly limited to detecting "hot Jupiter" exoplanets. Current large ground-based telescopes have primarily searched for planets around nearby stars using the radial velocity method, including work by the two 10 m diameter telescopes of the W.M. Keck Observatory, the 8 m telescope at the Gemini North Observatory, and the 3.6 m telescope at the European Southern Observatory (ESO). Networks, such as the Hungarian Automated Telescope Network (HATNet) and the Wide Angle Search for Planets (WASP), have been successful in detecting and characterizing planets through the use of a group of small, fully automated telescopes that observe transits. Smaller ground-based telescopes can also directly hunt for extrasolar planets with adaptive optics (to compensate for atmospheric distortion) and a coronagraph (to block direct starlight). The ESO Very Large Telescope (VLT) will have the capability to analyze the atmospheric composition of Earth-sized or super-Earth-sized exoplanets. The joint ESO/National Radio Astronomy Observatory (NRAO)/National Astronomical Observatory of Japan (NAOJ) Atacama Large Millimeter/submillimeter Array (ALMA) will be capable of observing young giant planets that form in protoplanetary disks as well as giant planets orbiting nearby stars. ALMA made its first observations in 2013.

Future space-based missions include the European Characterising ExOPlanets Satellite (CHEOPS) and Planetary Transits and Oscillations of stars (PLATO) missions, the JAXA Space Infrared Telescope for Cosmology and Astrophysics (SPICA), and NASA’s Transiting Exoplanet Survey Satellite (TESS) and JWST. CHEOPS will be launched in 2017 and will observe already-detected planets for which only the mass, but not the size, is known (both sets of information together allow for a characterization of these planets). Instead of a random exoplanet search, CHEOPS knows where and when to look for transits and is therefore expected to have a high success rate. PLATO has been selected as a medium-class ESA mission and is due to launch in 2025. The mission combines accurate transit measurements with astroseismicity, which allows for determination of the age of the host star. SPICA is an international, though JAXA-led, mission designed to further our understanding of the birth of galaxies, stars, and planets. This 3.5 m IR telescope (possibly based on the ESA mission Herschel), which could be launched as early as 2025, will characterize the atmospheres of exoplanets via transit spectroscopy. JWST is another international, space-based IR telescope, led by NASA, that is proposed to launch in 2018. This mission will study the formation of stars and planetary systems and enable transit spectroscopy, primarily at IR wavelengths, which will allow researchers to characterize exoplanets via identification of atmospheric gases, surface features, and production of maps of thermal emission (see Chapter 7 for more on this). Transits, though not transit spectroscopy, are the primary focus of NASA’s TESS mission, which will launch in 2017. TESS is an all-sky survey of transit exoplanets, the mission goals of which include the discovery of Earth-sized planets in the habitable zones of nearby small stars; the 2-year baseline mission, however, allows too little time to attempt detection of planets in the habitable zones of Sun-like stars. Future ground-based activities include the ESO 30–60 m European Extremely Large Telescope (ELT), which is prioritizing the measurement of atmospheric properties of exoplanets that are the size of Venus or larger.

Proposed space-based and ground-based missions are the subjects of recommendations by several astronomy decadal surveys. One common, high-priority science objective noted by decadal surveys for 2015–2025 is to seek and study nearby habitable planets. For example, the Astro2010 New Worlds, New Horizons in Astronomy and Astrophysics (Committee for a Decadal Survey of Astronomy and Astrophysics, 2010) report recommends a select few missions, and it ranks these proposed missions within one of three categories: large-scale space-based, medium-scale space-based, and small-scale space- and ground-based. The highest recommended proposed large-scale space-based initiative is the 1.5 m WFIRST.
WFIRST could begin to search for exoplanets as early as 2024 via microlensing and may also include a coronagraph for direct imaging of planets larger than Earth. The highest recommended medium-scale space-based mission is the New Worlds Technology Development Program, which will lay the technical and scientific foundation for a future mission (beyond 2020) to study nearby Earth-like planets.

NASA has established a series of science and technology definition teams (STDTs) to explore options to pursue this kind of observation. An STDT has already studied the potential for WFIRST to directly observe large planets. Two of the four flagship missions—LUVOIR and HabEx—now have STDTs; the major goal of these missions would be to detect and characterize potentially habitable worlds. For more on these exoplanet characterization missions, exoplanet characterization, and the search for life beyond Earth, see Chapter 7.

6.4. In what other ways do we study places we are unable to visit?

Planetary missions are expensive and present significant technical challenges, so astrobiologists also use terrestrial analogues of other planetary bodies to further our understanding of the potential for life beyond Earth. Some of these are laboratory environments, while others are places on Earth that share some of the present or past characteristics of planetary bodies. While no Earth-based site can fully match the environment of another planet, these studies provide essential information on the physical and chemical processes that take place in these environments, on the limits of life, and on how life adapts to changing environments. The analog community is discussing rubrics (Hipkin et al., 2013) to improve our ability to compare and contrast different analog sites and to combine results from multiple sites that have complementary strengths. A detailed discussion of analog sites is beyond the scope of this primer, but some examples are given in Table 6 and the associated Fig. 26. More thorough reviews of analog sites can be found in the works of Marion et al. (2003), Fairen et al. (2010), and Léveillé (2010); a Catalogue of Planetary Analogues is available from ESA (esamultimedia.esa.int/docs/gsp/The_Catalogue_of_Planetary_Analogues.pdf).

Chapter 7. What Are the Signs of Life (Biosignatures) That We Could Use to Look for Life beyond Earth?

A biosignature is defined as any characteristic element, molecule, substance, or feature that can be used as evidence of past or present life and is distinct from an abiogenic

<table>
<thead>
<tr>
<th>Planetary body</th>
<th>Analog site</th>
<th>Key field site characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mars</td>
<td>Arctic and Antarctic sites, including permafrost and ice surfaces in the Canadian High Arctic, Siberia, Svalbard, Norway, and the Antarctic dry valleys (Steven et al., 2006; Hart et al., 2011; Rull et al., 2011) Volcanic and hydrothermal environments, including Iceland, Hawaii (USA), Yellowstone National Park (USA), and hydrothermal vents (Jorge-Villar and Edwards, 2010) Atacama Desert, Chile; sand dunes and deserts, Morocco (Parro et al., 2011); hematite concretions in Utah canyone, USA (Chan et al., 2011)</td>
<td>Long-term exposure to extreme cold, extreme dryness, exposure to radiation Active and past volcanism and hydrothermal activity, on land, under water, or in the presence of ice. Presence of basalt (common rock type on Mars). Extreme dryness, presence of perchlorates (Atacama), environments dominated by wind-driven processes (Morocco), geological evidence of water-driven processes resulting in hematite concretions similar to the hematite “blueberry” concretions found on Mars (Utah) Acidic, iron-rich environment, similar to conditions believed to have existed on early Mars Subglacial lakes beneath an ice layer hundreds of meters to several kilometers thick. Believed to have been isolated from the surface for hundreds of thousands of years Mg-Na-SO₄-Cl brine proposed as an analogue for the europan ocean Sulfur deposits on icy surface Hydrocarbon-dominated liquid environment</td>
</tr>
<tr>
<td>Europa</td>
<td>Antarctic subglacial lakes, e.g., Lake Vostok (D’Elia et al., 2008; Jones, 2012), Lake Whillans (Priscu et al., 2013), and Lake Ellsworth (Siegert et al., 2007) Lake Tirez, Spain (Prieto-Ballesteros et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>Titan</td>
<td>Rancho La Brea Tar Pits, USA (Kim and Crowley, 2007)</td>
<td></td>
</tr>
</tbody>
</table>

Many other sites are also used (see reviews cited above).
background. In astrobiology, the terms *biosignature* and *biomarker* are sometimes used interchangeably, although biosignature is often preferred to avoid confusion with clinical biomarkers in medicine (*e.g.*, the presence of an antibody in a blood test is a biomarker that may indicate infection). In many cases, the unambiguous identification of biological activity can be very challenging.

In this chapter, we discuss controversial examples such as the martian meteorite ALH84001 and the Viking lander’s experimental results. The general issue, when selecting biosignatures to include in the search for life, is that targeting very specific molecules or other signs of terrestrial life may allow for the design of sensitive instruments, but it may also contribute to researchers making assumptions about the likely nature of extraterrestrial life. On the other hand, looking for more general kinds of evidence may yield ambiguous results that cannot unequivocally be attributed to life. Thus, the ongoing work of cataloguing indicators of

**FIG. 26.** Examples of some of the planetary analog sites described in Table 6. (A) Microbial growth (pink color) on glacial ice, Svalbard (photo credit: Katherine Wright). (B) Microbial mats in Yellowstone Hot Springs (photo credit: Katherine Wright). (C) Sand dunes, Morocco (photo credit: Katherine Wright). (D) Hematite concretion, Utah (photo credit: Katherine Wright). (E) Rio Tinto, Spain (photo credit: Dámnait Gleeson). (F) Elemental sulfur deposits on the surface of Borup Fiord Pass Glacier, Canada (photo credit: Katherine Wright).
both present and past life on Earth remains a primary goal in astrobiology; this biosignature work increases our understanding of the origin and evolution of life on Earth and helps in the search for life on other planets.

One of the first questions that comes to mind when designing missions to search for past and present life on other planets is, What are the indications of life here on Earth? On our planet, biosignatures exist at multiple scales. The study of biosignatures of terrestrial life that are identifiable in close-up is instructive for in situ (in place) investigations to be conducted on Mars and other bodies of our solar system. Instruments aboard lander probes are based on systems already in use for terrestrial applications and have been adapted to function reliably in the environmental conditions of a target planetary body. In the same way, the study of biosignatures that are identifiable by terrestrial remote sensing (from satellites or space probes) constitutes an important test bed for the remote recognition of biosignatures in distant objects such as extrasolar planets. Table 7 lists different potential biosignatures along with example instruments that can detect these various signs, as explained in this chapter.

7.1. What are the physical and chemical signs of life that can be detected in situ?

7.1.1. Visual biosignatures. 7.1.1.1. Active life. Earth’s simplest life-forms present common morphological characteristics that can be recognized visually. All known life is composed of internal volumes isolated from the surrounding environment by a cell membrane and sometimes an additional cell wall (see Chapter 4). These cellular structures can, in principle, be detected microscopically. However, the typical sizes of cells are at the lower limit that can be studied with light microscopes (e.g., a bacterium is usually on the order of 1 μm in length). Microbiologists look for a number of signs to determine whether a given cell-like structure is alive, including its size, shape, context (cells are sometimes found in groups or form chains of cells end to end), movement, and reproduction. Fluorescent dyes that bind to specific biomolecules are sometimes employed to make these kinds of determinations. However, none of these signs are completely definitive on their own, so a combination is required to evaluate the presence of extant life.

7.1.1.2. Fossils. Signs of ancient, long-dead life are harder to spot due to metabolic inactivity and cellular degradation. The most obvious and easily identified signs of ancient life on Earth are body fossils—imprints of the hard parts of animal skeletons and plants. While body fossils are one of the most robust signs of life (abiotic processes do not produce features as complex as a dinosaur skeleton or bi-valve shell), multicellular life is very unlikely elsewhere in the Solar System, and we will not be able to study extrasolar planets in situ for centuries.

Microscopic fossils (or microfossils) of the earliest unicellular organisms on Earth are much less clear-cut. There are heated debates on the oldest fossils of life on Earth. Several different controversial features have been found in the Apex chert from Western Australia, dated at ~3.5 Ga. Some researchers claim that these morphological features are the fossilized cells of filamentous bacteria (Schopf, 1993; Schopf et al., 2002), while others argue that they are abiogenic (Brasier et al., 2005).

| Table 7. Different Types of Potential In Situ and Remote Biosignatures |
|-------------------------|--------------------------|-----------------|----------------------------------|
| **Type**                | **Biosignatures**         | **Examples**    | **Appropriate techniques**      |
| In situ                 | Visual biosignatures      | Direct observation of active life | Cellular structures (possibly seen to be motile or reproducing) | Microscopy or macroscopic imagery |
|                        | Fossils                  | Fossilized cells | Protins or nucleic acid polymers (e.g., DNA, RNA) | Gas chromatography–mass spectrometry, Raman spectroscopy, X-ray spectroscopy |
|                        | Biological macromolecules| Stromatolites or endolithic microboringings | Breakdown products of biomolecules, such as hohanoids or steranes |
|                        | Molecular fossils        | Biogenic biases, such as isotopic fractionation or homochirality | Gradients of redox species in column of lake water |
|                        | Molecular evidence of metabolism | | |
|                        | Thermodynamic or kinetic disequilibrium within environment | | |
|                        | Biominerals              | Certain silicate, carbonate, or iron minerals, or metal enrichments of, e.g., Cu, Mo, Ni, W |
| Remote                 | Spectral biosignatures   | Large-scale environmental disequilibrium, e.g., O₂ and CH₄ | Atmospheric disequilibrium, | IR and visible spectroscopy |
| Spatially resolved     | Photosynthetic life      | Red edge of vegetation | Roads, cities, agriculture, large-scale landscape modification | Optical imaging |
| Electromagnetic        | Geometrical structures of intelligent life | Radio or optical signals from a civilization |
| emissions              | Intelligent broadcasts   | | Radio or optical sky surveys |

7.1.1.1. Active life. Earth’s simplest life-forms present common morphological characteristics that can be recognized visually. All known life is composed of internal volumes isolated from the surrounding environment by a cell membrane and sometimes an additional cell wall (see Chapter 4). These cellular structures can, in principle, be detected microscopically. However, the typical sizes of cells are at the lower limit that can be studied with light microscopes (e.g., a bacterium is usually on the order of 1 μm in length). Microbiologists look for a number of signs to determine whether a given cell-like structure is alive, including its size, shape, context (cells are sometimes found in groups or form chains of cells end to end), movement, and reproduction. Fluorescent dyes that bind to specific biomolecules are sometimes employed to make these kinds of determinations. However, none of these signs are completely definitive on their own, so a combination is required to evaluate the presence of extant life.

7.1.1.2. Fossils. Signs of ancient, long-dead life are harder to spot due to metabolic inactivity and cellular degradation. The most obvious and easily identified signs of ancient life on Earth are body fossils—imprints of the hard parts of animal skeletons and plants. While body fossils are one of the most robust signs of life (abiotic processes do not produce features as complex as a dinosaur skeleton or bi-valve shell), multicellular life is very unlikely elsewhere in the Solar System, and we will not be able to study extrasolar planets in situ for centuries.

Microscopic fossils (or microfossils) of the earliest unicellular organisms on Earth are much less clear-cut. There are heated debates on the oldest fossils of life on Earth. Several different controversial features have been found in the Apex chert from Western Australia, dated at ~3.5 Ga. Some researchers claim that these morphological features are the fossilized cells of filamentous bacteria (Schopf, 1993; Schopf et al., 2002), while others argue that they are abiogenic (Brasier et al., 2005).
7.1.1.3. Artifacts of life. Less controversial are ancient stromatolites, which date back to about 3.45 Ga (Walter et al., 1980; Allwood et al., 2009). Stromatolites are large layered structures of accreted sedimentary grains that are formed by microbial mat communities (see Fig. 27), although abiotic processes could form structures that mimic stromatolites (Grotzinger and Rothman, 1996). The oldest known potential eukaryotic microfossils are the 1.87 Ga impressions of filaments similar to coiled green algae (Knoll et al., 2006).

In addition to the preserved remains of organisms themselves, there are many different types of trace fossils, which record the behavior or activities of organisms. These typically consist of morphological features such as burrows, borings, footprints, feeding marks, and root cavities. Mineralized microscopic tubules, called endolithic borings, are a possible biosignature that can be found within rocks. Examples of such features include those found in the Barberton greenstone belt in South Africa (3.5–3.4 Ga) and the Pilbara Craton in Western Australia (3.35 Ga). These could potentially be the result of microbial activity (Banerjee et al., 2006; Fliegel et al., 2010); if so, these would be among the earliest morphological signs of life of Earth. It is also possible, however, that the mineralized tubules observed in the rocks of both these formations may have had an abiotic origin; so, to date, no clear determination has been made as to whether these features are definitive biosignatures or structures that resulted from abiogenic processes. Since microfossils and trace fossils are often difficult to interpret on their own, arguments for biogenicity are much better founded with additional, complementary lines of evidence.

Visual biosignatures are potentially the easiest for a robotic probe operating in an extraterrestrial environment to spot. Macroscopic signs of life, such as fossilized stromatolites, may be detected visually, and microscopic imagers could reveal fossilized bacteria or endolithic borings. Cameras are crucial components of the scientific payload of landing probes, as they provide a general overview of the landscape and the context for other measurements and analyses, as well as allowing for rover navigation and hazard avoidance. Use of a range of filters enables multispectral imaging that can help determine the mineralogy at a site. Paired cameras yield stereoscopic imaging and allow reconstruction of the three-dimensional lay of the landscape and range to particular targets. All of these are essential for understanding the context of any observations of astrobiological significance.

7.1.2. Chemical biosignatures. 7.1.2.1. Biological macromolecules and molecular fossils. On a smaller scale, biomolecules that make up terrestrial living cells (and by extension possibly extraterrestrial cells), and especially biological macromolecules—carbohydrates, lipids, proteins, and nucleic acids (including RNA and DNA)—could in principle be used to infer that life once existed in a particular environment. Unfortunately, most common biomolecules are easily thermally or chemically modified and degraded, so the macromolecules of life tend to be poorly preserved. However, the chemical breakdown (diagenesis) of some biomolecules produces recognizable compounds that are preserved long-term. These breakdown products are often dubbed “molecular fossils” (Brocks et al., 1999; Summons et al., 1999).

FIG. 27. Stromatolites are still present today in a few places on Earth, such as Shark Bay in Western Australia (left). These modern stromatolites are very useful for understanding structures that were formed much longer ago (right). [Credits: Richard Arculus (left) and NASA Ames (right)]
Of the four main classes of biomolecules, the lipids, which are found in all cellular membranes and used for energy storage, are best preserved. Table 8 lists some example lipid biomolecules and their molecular fossils. Some molecular fossils are claimed to be indicative of broad groups of organisms or even a single group of organisms. So specific molecular fossils detected within host rocks on Earth have been used to suggest dates for the first emergence of particular groups of organisms or infer the ancient diversity of organisms present in the local ecosystem (Eigenbrode, 2008). However, these claims must always be subject to the caveat that the same lipid could be produced by other organisms yet to be identified. For example, although 2-methyl hopanes are known to be the breakdown product of 2-methylbacteriohopanepolyols, which were once thought to be produced only by cyanobacteria, it has recently been shown that 2-methylbacteriohopanepolyols are produced by other bacteria as well (Rashby et al., 2007; Welander et al., 2012).

These particular compounds could, however, be overly specific to terrestrial life. So it may be preferable to search for broader characteristics of life, such as evidence that organic molecules have been synthesized by biological processes.

7.1.2.2. Molecular evidence of metabolism. 7.1.2.2.1. Chirality. Many organic molecules are chiral, which is to say they can exist in two different mirror-image versions, called optical isomers or enantiomers (Fig. 28). A sample of such molecules is said to be a racemic mixture if it contains equal proportions of both enantiomers. Biological organic molecules, however, exhibit homochirality, which means they show domination of one or other of the two enantiomers. For example, almost all amino acids used by life on Earth are said to be “left-handed” (L), while almost all sugars are “right-handed” (D). Finding a sample of organic molecules to exhibit homochirality would thus be a strong indicator that they have been synthesized biologically (although organics do racemize over time after death of the organisms). The reason terrestrial life employs l-amino acids and d-sugars is not clear (Wallace, 2011), but once one enantiomer becomes dominant in prebiotic chemistry, the selection seems to become locked in.

7.1.2.2. Isotopic fractionation. Another signature of the action of metabolism is manifested in the ratio of two stable isotopes. Isotopes are variants of the same element with different atomic masses because they possess a different number of neutrons in the nucleus. Stable isotopes do not undergo radioactive decay on geological/astronomical timescales. Biological action causes disequilibrium in the isotope ratios of stable elements because enzymatic processes preferentially use lighter isotopes, resulting in the heavier ones accumulating in the surrounding environment—the isotopes become separated or “fractionated.” Analysis of the two stable carbon isotopes, $^{13}\text{C}$ and $^{12}\text{C}$, is the most common isotopic system used as a biological indicator. The detection of a biological isotopic signature therefore requires the accurate measurement of the isotope ratio in more than one reservoir—organic carbon in a lake-bed sediment and dissolved carbon dioxide in the overlying water column, for example—as well as knowledge of the original ratio to determine whether fractionation has occurred.

The degree of fractionation is commonly expressed by using the delta notation, which indicates the shift in isotope ratio from a given standard. For example, fractionation of the carbon isotopes $^{13}\text{C}$ and $^{12}\text{C}$ (with six and seven neutrons in the nucleus, respectively) is given as $\delta^{13}\text{C}$ ($\%$), which is defined as the ratio of $^{13}\text{C}$ to $^{12}\text{C}$ in the sample divided by the $^{13}\text{C}$:$^{12}\text{C}$ ratio in the standard, and multiplied by 1000 to yield a per mil measure. The preferential fixation of $^{12}\text{C}$ into biological molecules yields a negative $\delta^{13}\text{C}$ in biologically synthesized (biogenic) organics and a corresponding increase in $\delta^{13}\text{C}$ in the local microneche or wider environment. The ratios of other stable isotopes are also analyzed, including oxygen, $\delta^{18}\text{O}$ ($^{18}\text{O}/^{16}\text{O}$); nitrogen, $\delta^{15}\text{N}$ ($^{15}\text{N}/^{14}\text{N}$); sulfur, $\delta^{34}\text{S}$ ($^{34}\text{S}/^{32}\text{S}$); and iron, $\delta^{56}\text{Fe}$ ($^{56}\text{Fe}/^{54}\text{Fe}$).

Ruling out abiotic fractionations is one of the greatest challenges in using isotopic fractionation as a biosignature. Abiotic processes such as serpentinization can also alter

<table>
<thead>
<tr>
<th>Table 8. Example Lipid Biomolecules and Their Molecular Fossils</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biomolecule → molecular fossil</strong></td>
</tr>
<tr>
<td>Isoprenoids → branched alkanes</td>
</tr>
<tr>
<td>Sterols → steranes,</td>
</tr>
<tr>
<td>aromatic steroids</td>
</tr>
</tbody>
</table>

FIG. 28. Many biological molecules, like amino acids, are chiral; they exist in two mirror-image versions (just like our left and right hand), but biochemistry uses only one enantiomer. (Credit: NASA)
isotope ratios (Rothschild and Des Marais, 1989), so abiotic processes must be ruled out before biology can be identified as the source of fractionation. Such discrimination can occur based on the size of the fractionation or by detecting isotopic fractionations of multiple elements. The potential confusion and controversy caused by these abiotic processes is exemplified by some of the oldest (but disputed) evidence for life on Earth. The oldest proposed chemical signs of life on Earth are the strongly negative $\delta^{13}C$ values analyzed from carbonaceous inclusions in altered banded iron formations from Akilia Island, West Greenland, $\sim 3.85$ Ga (Mojzsis et al., 1996; Mojzsis and Harrison, 2000). However, these claims are controversial (van Zuilen et al., 2002; Lepland et al., 2005) because the cycling of hydrothermal fluids may produce similar carbon isotope ratios.

7.1.2.3. Chemical disequilibria. A potential biosignature of living processes that are still active is chemical disequilibrium in the surrounding environment. Living cells maintain an internal cellular environment that is different from the external environment (cellular homeostasis). Energy must be extracted from the environment to support metabolic processes like maintenance, growth, and reproduction. Cells obtain energy from chemical reactions or from light. Chemical reactions must be thermodynamically favorable to support life (see Chapter 5 for more detail) but may be slow (kinetically unfavorable) when the reaction proceeds without life. Cells use enzyme catalysts to accelerate reaction rates so that the reaction occurs faster than it would under abiotic circumstances. Photosynthetic cells can use light energy to drive a reaction that is, otherwise, energetically unfavorable, and this may create a thermodynamic chemical disequilibrium (e.g., the production of oxygen in oxygenic photosynthesis). Thus, an amount of a species or a ratio between two species that can only be explained by thermodynamically or kinetically unfavorable reactions may possibly be an indicator of life, although abiotic processes (not involving life) might also be responsible. Such a biosignature of active life can be detected in situ, for example in the measured concentration gradients of redox species in a column of lake water which would rapidly dissipate without life (Nealson, 1997). Section 7.2.2 covers how planetary disequilibria may be detected remotely.

7.1.2.4. Biominerals. Another record of the action of life, which persists far longer than chemical disequilibrium in the environment, is the diverse and widespread phenomenon of biomineralization. Biogenic minerals, or biominerals, are those that are formed either directly or indirectly from living organisms. Some biominerals crystalize within living organisms as a result of cellular activity, while others are formed extracellularly. Deposits of calcium carbonate, calcium phosphate, sulfur, manganese, and iron oxides have all been attributed to biological activity (Knoll, 2003). However, it can be very difficult to distinguish between minerals that are produced specifically by microbial metabolic processes and minerals produced by purely abiotic processes. In 1996, McKay et al. identified carbon-rich morphological features in the martian meteorite ALH84001 (Fig. 29, left) that looked very similar to cells (Fig. 29, right). Their claims were initially supported by the presence of highly ordered magnetite crystals, which were thought at the time to be only produced by magnetotactic bacteria (Thomas-Keprta et al., 2000). (Magnetotactic bacteria create tiny magnets inside their cells for navigation/orientation purposes.) Shortly after that historic announcement (http://www2.jpl.nasa.gov/snc/clinton.html), the findings were disputed by a number of studies that showed abiotic production of similar small, highly ordered magnetite crystals (Golden et al., 2001, 2004). Because such studies have demonstrated abiotic processes that could lead to the features in ALH84001, the original conclusion of biosignatures in that sample remains controversial.

7.1.3. What techniques do we use to detect life in situ? Alongside an appreciation of the ways in which life modifies its environment or what features it leaves behind, an equally critical facet of biosignatures in astrobiology is the operation of instrumentation that is able to detect these signs.
7.1.3.1. Mass spectrometry. Mass spectrometry is a powerful analytical technique that has the capability to detect and identify organic compounds and is sensitive to the isotopic composition. This process involves first ionizing the sample (usually by stripping electrons from it with an energy source) and then separating the resulting stream of charged molecules by their mass-to-charge ratio. This produces a mass spectrum that must be interpreted to reveal details of the chemical composition. Available mass spectrometers use a variety of methods to create and then sort the ions. For example, the Sample Analysis at Mars (SAM; Mahaffy, 2009) instrument aboard the MSL rover (“Curiosity”) employs a quadrupole mass spectrometer with an electric field applied to four parallel metal cylinders to deflect the trajectories of ions in order to sort them.

7.1.3.2. Gas chromatography–mass spectrometry. A mass spectrometer (MS) is often used in conjunction with a gas chromatograph (GC), which separates compounds by their interactions with the material in the GC column and thus how long they take to pass through. The stream from the GC is passed immediately into a MS to produce a combined GC-MS instrument with better identification ability than either technique alone. Sensitive gas chromatography–mass spectrometry can also analyze the stable isotope ratio of specific compounds. It is for these reasons that gas chromatography–mass spectrometry is commonly used on planetary missions, including the Viking Mars landers and MSL. Mars Science Laboratory’s SAM package (Mahaffy, 2009) includes a tunable laser spectrometer (TLS) that can determine the isotope ratios for carbon and oxygen in carbon dioxide. The TLS is also capable of measuring trace levels of methane and its isotopic composition and can detect methyl mercaptan, another biogenic gas (Vance et al., 2011).

7.1.3.3. Raman spectroscopy. Raman spectroscopy is another technique suited to astrobiological exploration. The target is illuminated with monowavelength laser light, and the scattered light is collected and spectroscopically analyzed. Wavelength shifts caused by inelastic scattering of the laser light reveal information on the vibrational and rotational modes within the sample, which can be used, for example, to identify organic molecules within the sample. Raman spectroscopy has proved to be very successful in detecting the complex biomolecules of microbial colonies in the most hostile environments on Earth, and has the advantage of no requirement for reagents or sample preparation, although Raman scattering is a photon-inefficient process (Ellery and Wynn-Williams, 2003; Jorge Villar and Edwards, 2006; Tarcea et al., 2008; Marshall et al., 2010). Raman spectroscopy is sensitive (although the signal depends on the illumination wavelength) not only to microbial pigments, such as chlorophyll and protective carotenoids, but also to mineralogy, and so also provides information on the geochemistry of the microbial habitat. ExoMars and Mars2020 will be the first planetary probes to deploy Raman spectroscopy, the former using it to analyze subsurface samples of martian soil retrieved by the rover’s drill system (Rull et al., 2010).

7.1.3.4. Immunoassay techniques. Immunoassays are another method proposed for detecting and identifying low concentrations of organics based on recognition proteins known as antibodies. These are molecule-specific detection techniques that would present us with strong biosignatures but that can only detect life that has a similar (or identical) chemical composition as Earth-based life. The Life Marker Chip (LMC; Sims et al., 2005) and Life Detector Chip (LDChip300; Parro et al., 2011) are similar systems that use antibodies as molecular receptors to bind to, and thus detect, specific organic molecules or classes of molecules with very high specificity and sensitivity.

7.1.3.5. Fluorescence. The principle of fluorescence underlies a range of methods used terrestrially to study the chemical composition and biological activity of samples and is currently being extended to astrobiological applications. Fluorescence occurs when short-wavelength light (e.g., UV) is absorbed by organic molecules, which causes electronic excitation and subsequent emission of longer-wavelength light (e.g., visible) when the state relaxes. The optimum excitation and emission wavelengths depend on the molecular structure and its electronic environment, so fluorescence spectroscopy can not only detect fluorophores but also aid in their identification.

Prebiotic compounds such as polycyclic aromatic hydrocarbons (PAHs), biogenic compounds like tryptophan (an amino acid), and biological molecules like chlorophyll all contain aromatic ring structures that render them intrinsically fluorescent under UV illumination. Such auto-fluorescence may offer a means by which to triage samples for the presence of organics before more discriminatory, but resource-intensive, techniques are used (Storrie-Lombardi et al., 2008). If a target compound is not itself fluorescent, it can be labeled with the addition of a specific compound that does fluoresce efficiently—a fluorophore tag. The autofluorescence of organic molecules and whole microorganisms, as well as applicable fluorophore dyes, is being investigated for use in biosignature detection instruments (Nadeau et al., 2008; Weinstein et al., 2008; Storrie-Lombardi et al., 2009; Benhabib et al., 2010; Dartnell et al., 2010).

7.1.3.6. Search considerations. A major consideration for the design and operation of instrumentation employed in the search for organic molecules, such as GC-MS or immunoassay-based equipment, is the key targets of interest—specific molecules that are the most robust biosignatures of past or present life—and their prioritization. Parnell et al. (2007) reported on the conclusions from an ESA workshop, listing the primary target molecules for biosignatures of extant organisms, molecular fossils, organics of meteoritic origin, and possible contaminants such as hydrazine rocket fuel. They stress the importance of ground-based experiments for assay development and the need to detect not just a single target but to assess the distribution of compounds.

Another critical factor in the detectability of potential biosignatures is how resilient they might be in their environment. For example, the persistence of possible biosignatures on Mars, both exposed on the surface and within the top few meters of the near-subsurface that are accessible with drills, is likely to be limited by the current hostile environmental conditions. On the martian surface, the unshielded flux of solar UV light rapidly photodegrades exposed organic molecules (ten Kate et al., 2005; Shkrob...
et al., 2010; Dartnell et al., 2012), the strongly oxidizing conditions of the wind-blown dust layer would chemically attack molecular biosignatures (Benner et al., 2000), and the ionizing radiation of cosmic rays penetrates several meters beneath the surface (Pavlov et al., 2002; Dartnell et al., 2007).

7.1.3.7. Case study: Viking landers. The Viking landers represent an important case study in astrobiology. Three experiments were included in the Viking biology package to test for metabolic activity of microorganisms in samples of the surface. The gas exchange experiment was designed to measure the uptake of carbon dioxide and the production of biologically relevant gases such as methane, hydrogen, and oxygen. The pyrolytic release experiment (wherein organics are broken down for analysis by heat) tested for carbon assimilation: the incorporation of radioactive $^{14}$C into organics. The labeled release experiment tested for the reverse process, the evolution of radio-labeled carbon dioxide gas after the addition of a nutrient soup of organic compounds containing $^{14}$C. Curiously, the labeled release experiment did indeed yield results consistent with biological activity: $^{14}$C-labeled gas was produced from the surface samples, but not when they had been sterilized by heating to 160°C (the negative control). However, the other two metabolism-detecting experiments could find no evidence of biological activity down to their limit of detectability. Furthermore, the GC-MS instrument could detect no trace of organic molecules in the samples. Interpretation of these contradictory results has been a matter of debate for over three decades (Klein et al., 1976; Klein, 1977, 1979; Ponnanperuma et al., 1977; Owen, 1979; Navarro-González et al., 2006, 2010; Biemann, 2007). More recently, the Phoenix lander detected perchlorates in the martian surface, which could explain the nondetection of organics even if present (Navarro-González et al., 2010). Perchlorate is highly oxidizing at the elevated temperatures of pyrolytic experiments and so may have destroyed martian organics before detection. In an effort to identify elusive organics and protect them from oxidation, MSL will attempt to solvent-extract organic compounds from surface samples before pyrolysis. For a recent review of the question of organics on Mars and the interpretation of the Viking results, see the work of ten Kate (2010).

7.2. How can we search for life remotely?

Remote sensing with Earth-based or space telescopes is always the first step to assess a planetary environment for habitability and potential biosignatures. For objects inside our solar system, remote observations precede measurements by instruments mounted on space probes (flyby spacecraft, orbiters) or delivered to the surface (landers, rovers). Due to the great distances to extrasolar planets (see Section 6.3), remote observations are the only method we have for exploring these objects. Within the confines of our solar system, remote observations offer resolution on features such as clouds, landforms, and liquid reservoirs. For objects outside the Solar System, spatial resolution will not be achievable in the next generation of space-based telescopes, which will not be large enough to resolve these features. Instead, near-term observations will rely on spectroscopy—wavelength-dependent observations—to obtain certain information. Temporally resolved spectra can be used to obtain latitudinal information on a planet (Cowan et al., 2009).

7.2.1. Visible and infrared spectroscopy. Every molecular species has unique, discrete energy transitions that correspond to absorption or emission of photons with different wavelengths because all molecules have a variety of energy states that are a function of the vibrational modes and configuration of electrons within the molecule. The energy needed for positive energy transitions can come from absorption of photons; conversely, the excess energy in negative energy transitions can go into emission of photons. In either case, the energy of the photon and, thus, its wavelength (another metric commonly used is the wavenumber—the reciprocal of the wavelength) must correspond to the energy gained or lost in the transition.

Spectroscopy can reveal the presence of specific molecules by identifying their characteristic energy transitions—a “dip” or a “peak” in the spectrum caused by molecular absorption or emission at molecule-specific photon wavelengths. For example, the reflection spectrum of Earth (Fig. 30) contains a series of clear dips in the spectrum between 0.8 and 3 mm. These dips are at wavelengths and resulting photon energies known to correlate with energy transitions in H$_2$O molecules; their presence is thus an indicator of the presence of H$_2$O in Earth’s atmosphere.

7.2.2. Atmospheric disequilibrium. Life has caused changes to the composition of Earth’s atmosphere (Lovelock, 1965) that are detectable by spectroscopy. Observations of these features include measurements made by spacecraft (Sagan et al., 1993; Grinspoon et al., 2008; Robinson et al., 2011) and by ground observations of “Earthshine,” light that is reflected off the Earth and then off the “dark side” of the Moon (Arnold, 2008). In all these observations, the most robust spectral evidence of life is the simultaneous presence of an oxidizing species, such as molecular oxygen (O$_2$), and a reducing species, like methane (CH$_4$). The mixture of oxygen and methane is far from “atmospheric equilibrium,” and on short timescales the two species—oxidant and reductant—react to produce carbon dioxide. The short lifetimes of these gases mean that they have to be constantly replenished at a fast rate in order for them to be maintained at detectable concentrations. In other words, their presence implies large sources of these gases. The most sustainable source of these gases at the required rates is life. However, disequilibrium is also caused by nonbiological processes, such as photolysis and physical mixing. The “disequilibrium” in this case is a change with respect to the state of the atmosphere without biological inputs.

Taking the atmospheric context into consideration is critical when determining whether the presence of a gas is due to biology. For example, oxygen and methane are not unambiguous biosignatures on exoplanets, especially when detected in the absence of the other. A planet on the inner edge of the habitable zone (see Chapter 5) could conceivably have a large abiotic buildup of oxygen through a runaway greenhouse effect while the oceans evaporate; after photodissociation of water in the atmosphere, hydrogen could escape, leaving oxygen behind (Kasting, 1988). Recently, several groups have studied the possibility of
detectable levels of oxygen and ozone formation in CO₂-rich planets orbiting M dwarfs (Hu et al., 2012; Tian et al., 2014), FGKM stars with low-H₂/high-CO₂ atmospheres (Domagal-Goldman et al., 2014), and pre-main sequence M stars (Luger and Barnes, 2015). Wordsworth and Pierre-humbert (2014) investigated oxygen and ozone formation in CO₂-rich low-pressure atmospheres.

Methane also is produced abiotically in hydrothermal systems. In certain planetary environments with high UV fluxes and abundant CO₂, it is possible to simultaneously detect abiotically produced O₃ and CH₄ in the UV-visible, making further contextualization of the planet vital to eliminate false-positive detections of life (Domagal-Goldman et al., 2014). On top of the biogenicity issues, other contextual issues could affect biosignatures. The detectability of Earth’s biomarkers would have been different in some earlier geological time periods (Kaltenegger et al., 2007), and the current composition of biomarkers would be different if the Sun had a different stellar energy distribution (i.e., color spectrum) (Selsis, 2000; Segura et al., 2003, 2005; Grenfell et al., 2007; Rugheimer et al., 2013). Due to effects such as these, it is critical that atmospheric spectra are interpreted with the aid of atmospheric models and with knowledge of the stellar UV radiation environment of the host star (Rugheimer et al., 2015). These models can simulate the context for the presence of a potentially biogenic species, better constrain its concentrations, and determine whether the source of that species could instead be abiotic.

Gases other than O₂ and CH₄ have also been proposed as biomarkers whose detectable presence in an atmosphere implies life. Nitrous dioxide (N₂O) is primarily biogenic on Earth and does not have a significant abiotic production mechanism. It has been proposed that ozone (O₃) and ethane (C₂H₆), the photochemical by-products of O₂ and CH₄, respectively, could help determine the presence and amount of O₂ and CH₄. Ozone is detectable at very low concentrations; therefore it is a very good indicator for even trace O₂ concentrations; therefore it is a very good indicator for even trace O₂ concentrations (Léger et al., 1993; Segura et al., 2003). Ethane could be indicative of a biological flux of organic compounds into the atmosphere and could be useful in determining the biogenicity of a methane signature (Domagal-Goldman et al., 2011). Many gases

---

**FIG. 30.** The spectra of Earth and the Sun. In the visible portion of the spectrum, indicated by the rainbow, light from the planet is simply reflected sunlight, whereas the emission spectrum of the warm planet itself can be observed in the IR. Observing in the IR also offers a far more favorable contrast ratio between the intensity of the planetary emission and that of its star (10⁶) compared to visible wavelengths (10¹⁰). The spatial resolution of the terrestrial spectrum shown is similar to that being considered for TPF-C and Darwin missions. (Credit: Tyler Robinson)
are produced by life on Earth, including NO₂, NH₃, CH₂Cl, and CH₃SH, but have such low abundances in the atmosphere that they would not be detectable across interstellar distances using currently proposed technologies (Meadows and Seager, 2010). Other species, such as water and/or carbon dioxide, are produced by life but have significant abiotic sources and so are not strong biosignatures. (However, their detection does have implications for the habitability of the planet; see Chapter 6.)

The discovery of methane in the atmosphere of Mars (Formisano et al., 2004; Krasnopolsky et al., 2004; Mumma et al., 2004, 2009) provides an example of an atmospheric disequilibrium that has been detected by remote spectroscopy. Methane should have a short lifetime on Mars, on the scale of hundreds of years, implying that there is a recent or current source on Mars. The two possibilities for this source are intriguing to astrobiologists: geological activity and liquid water in the subsurface, or subsurface biology. Mumma et al. (2009) also claimed to have discovered temporal and spatial variability in martian methane. However, this claim is controversial. Such variations are difficult to account for chemically (Zahnle et al., 2011), and the MSL SAM instrument has been unable to detect atmospheric methane to an upper limit of only 1.3 ppbv (Webster et al., 2013).

7.2.3. Photosynthetic pigments. Certain surface-based biosignatures may also be observable in the spectrum of a planet. On Earth, the pigments in photosynthetic vegetation display a broadband spectral feature, the vegetation red edge (VRE), which is a sharp increase in the reflectance at ~700 μm (Seager et al., 2005; Arnold, 2008). This characteristic could, in principle, be detected as a few percent increase in the nonresolved reflection spectrum of the planet and thus constitute evidence of life (Hamdani et al., 2006). In practice, however, it is very difficult to detect this faint VRE in Earth’s spectra, as it is easily masked by clouds. On planets around stars with a different spectral energy distribution (color spectrum) from the Sun, photosynthetic organisms may create features analogous to the VRE but at other wavelengths (Kiang et al., 2007a, 2007b; Kiang, 2008). A confounding factor for such a potential biosignature is that similar features can be created by minerals. For example, sulfur and cinnabar are known to create abrupt reflectance edges that are similar both in shape and in strength to the VRE.

7.2.4. Transit spectroscopy. A technique currently in use for analyzing the spectra of planets is transit spectroscopy. When a planet passes in front of its star, the amount of light blocked depends on the absorption of its atmosphere, which varies with wavelength. A transmission spectrum of the planet is obtained by monitoring the effective diameter at various wavelengths. This technique has been used to identify sodium and water in the atmosphere of extrasolar gas giants (Charbonneau et al., 2002; Kreidberg et al., 2014b). This method has also been used to investigate the bulk composition and to detect the presence of clouds in the atmosphere of super-Earth GJ 1416 b (2.7 REarth, 6.5 MEarth; Berta et al., 2012; Kreidberg et al., 2014a). Further exoplanet spectroscopic characterization will be possible with large ground-based observatories with high-resolution spectroscopy (Snellen et al., 2015) and with the launch of JWST (Clampin et al., 2009; Doyon and Fullerton, 2012) or dedicated spacecraft like CHEOPS (Broeg et al., 2013). The transmission spectra of smaller, potentially habitable extrasolar planets has been modeled (Deming et al., 2009; Kaltenegger and Traub, 2009) and tested by measuring the reflected light from Earth during a lunar eclipse (Pallé et al., 2009). These studies indicate that future missions may allow the identification of the strongest features in an Earth-like atmosphere (e.g., ozone, oxygen, carbon dioxide, water, and potentially methane). For example, JWST could identify ozone or water in the atmosphere of a planet orbiting one of the nearest low-mass stars, although it would require hundreds of hours of observation to do so.

7.2.5. Characterization of Earth-like planets. The current plans to characterize Earth-sized planets rely on blocking out the light of the parent star in order to take a spectrum of the planet alone. There are three major mission architectures that have been proposed to accomplish this: (1) an external occultor or starshade that physically blocks the starlight; (2) a coronagraph that cancels out the starlight with advanced internal optics; and (3) an interferometer, which cancels the starlight through interference of light from multiple telescopes. The first two missions are both designed to observe in the visible to near-IR (0.5–2 mm; see Fig. 30), a region of the spectrum that could lead to observations of H₂O, O₃, O₂, CO₂, and CH₄ and surface vegetation. The latter mission (an interferometer) would operate in the mid-IR (~6 to ~20 mm, Fig. 30), where CO₂, H₂O, O₂, CH₄, NH₃, and N₂O could be observed. None of these concepts are being planned or considered for flight in the current decade. However, there is significant effort to further develop these mission concepts for flagship opportunities in future decades.

7.2.6. What about intelligent life? Most of the above biosignatures are produced by single-celled life that existed since the beginning of life on Earth. But in the very recent history of Earth, our civilization has been producing additional signs of life.

Notably, since the advent of wireless communication in the 1930s, humans have been sending electromagnetic signals that can be observed across interstellar space. One way to conduct a search for extraterrestrial intelligence (SETI) is thus to search for similar transmissions from other civilizations. That was done through numerous campaigns all across the electromagnetic spectrum, principally in radio [e.g., the Center for SETI Research uses the Allen Telescope Array (ATA); www.seti.org or the UC Berkeley/Cornell SERENDIP project using Arecibo Observatory], in the microwave (NASA’s Microwave Observation Program and the SETI Institute’s Project Phoenix; Tarter, 2001), and in the optical (OSETI project led by Harvard and Princeton Universities; Howard et al., 2004).

The best candidate for an extraterrestrial radio broadcast is a single, very strong, narrowband signal detected in 1977. It was dubbed the “Wow!” signal, after the reaction of the observer who first saw it at the Ohio State radio observatory. Possible sources of false positives, including satellites, weather balloons, and radio interference, have been excluded, and the origin of the signal is still a mystery today. However, this potentially interesting observation has not been confirmed or repeated (Kraus, 1979).
Besides this signal, there have been no promising detections of extraterrestrial intelligence (ETI) emissions since the beginning of active SETI research in the 1960s (Cocconi and Morrison, 1959; Drake, 2008). However, given the huge variety of wavelength and directions in which ETI could communicate, the possibility of ETI cannot be ruled out. It would take very significant resources to fully explore all these possible forms of emission, but this search will be narrowed with guidance from the detection of potentially habitable planets (see Chapter 6) or other biosignatures (see Section 7.2).

Another potential route for detecting signs of intelligent life may be the spectral features in extrasolar planet atmospheres of “artificial” molecules that could only be produced by an industrial civilization. Examples of such molecules include chlorofluorocarbons (CFCs) and other pollutants created by humanity that can be measured by spacecraft orbiting Earth (Carrigan, 2010). However, the detection of these molecules across interstellar space would require atmospheric accumulation orders of magnitude greater than that on Earth, or spectral resolution and sensitivity well beyond any near-term detection techniques (Beichman et al., 1999).

Other more speculative ways of detecting ETI have been proposed. One would be to search for the IR emission signal from a Dyson sphere, a gigantic structure built around a star by an advanced civilization to collect energy. Such proposals, however, are largely based on “science fiction technologies,” as we do not know how to engineer such a construction ourselves, and it may not even be possible. In October 2015, the Kepler space telescope observed an unusual signal in its data; the star KIC 8462852 showed irregularly shaped dips in its flux, dropping below 20%. This finding implies large amounts of opaque material must be orbiting the star to account for the large dip in brightness. The leading explanation for this anomaly is the chance observing of a swarm of dusty comet fragments in orbit around the star. However, it has been speculated that the data could represent a Dyson swarm (a variant of a Dyson sphere); this possibility garnered tremendous media attention. Follow-up SETI observations to look for signs of intelligent life have yet to produce signs of radio technology (Boyajian et al., 2016).

Chapter 8. What Relevance Does Astrobiology Have to the Future of Life on This Planet?

The rapid increase of human technical capabilities and knowledge of the nature of our solar system and our local neighborhood in the Galaxy has led to increased attention on the possibility of finding life on other planets. For the first time in human history, it is possible to apply the scientific method to investigate the possibility that life exists elsewhere in the Universe. This question is directly linked to the question of the origin and evolution of life on Earth, as well as the future of humankind. This chapter provides a very brief introduction to some of the ways in which astrobiological research has wider relevance to our everyday lives.

8.1. How can astrobiology contribute to solving real-world problems?

Pure research (i.e., research carried out without a specific practical application in mind at the time the research was performed) has often resulted in significant breakthroughs that have subsequently allowed the development of practical applications that are useful to society. In general, scientists support the need for pure research even when there is no idea how that research will eventually prove beneficial.

As described in Chapter 1, astrobiology includes a wide range of issues. NASA has defined astrobiology as being “the study of the origin, evolution, distribution, and future of life in the universe” (NASA Astrobiology website: http://astrobiology.nasa.gov). One of the earliest examples of how astrobiologically relevant pure research has subsequently led to practical applications is the use of astronomy for navigation throughout history (Lewis, 1994). A more recent example is the effect of atmospheric chlorine on ozone concentrations. Research demonstrating the critical role chlorine played in the atmospheric chemistry of Venus (Prinn, 1971; McElroy et al., 1973) preceded our understanding of chlorine’s critical role in the depletion of stratospheric ozone in Earth’s atmosphere (Molina and Rowland, 1974; Stolarski and Cicerone, 1974). This latter realization eventually led to the Montreal Protocol, which has substantially reduced anthropogenic emission of chlorine-containing species. The improved understanding of Earth obtained through observations of Venus is an example of the value of comparative planetology. We can ground-truth models and tools designed for Earth by applying them to other planets and improve them when we fold in the knowledge we gain when studying our neighbors. Such lessons have so far been limited to the physical systems in a planet, but if and when we discover life beyond Earth, these lessons will expand to the living planet and its interactions with the planet’s physical systems.

The desire to explore space has led to the creation of satellites that have enabled us to better track and understand weather, climate, and land surface changes that directly affect people’s lives (JAXA, 2008; NASA Science, 2014; ESA, n.d.). In geology, the study of plate tectonics has led to an improved ability to understand where earthquakes, volcanic eruptions, and tsunamis are likely to occur, offering the possibility that people living in those areas may be able to prepare better for such events. In biology, the discovery of rules that govern inheritance in pea plants by Mendel (Blumberg, 1997) began the study of genetic inheritance, which has revolutionized our understanding of genetically inherited diseases and improved our ability to detect and treat such diseases. Exploration of the hot springs at Yellowstone led to the discovery of the Taq DNA polymerase from Thermus aquaticus (Freeze and Brock, 1970; Gelfand et al., 1989; Brock, 1997), which was subsequently used to develop the polymerase chain reaction (PCR) technique (Mullis et al., 1986). PCR is a laboratory method used to amplify the number of fragments from a DNA template for different applications, including molecular identification, disease diagnostics, paternity tests, and forensic science in criminal investigations. PCR has completely revolutionized the entire field of molecular biology, impacting directly on medical research, and was an essential part of the process used to sequence the human genome.

Examples from research specifically focused on the search for extraterrestrial life include the fact that the search for extinct and/or extant life on Mars is forcing the miniaturization of equipment, especially spectrometers (Holland
et al., 2003; Jehlicka et al., 2010). These advancements may leverage or lead to improvements in biomedical analyses or breakthroughs for industrial processes. For example, Raman spectroscopy, a laboratory technique used to observe vibrational, rotational, and other low-frequency changes in molecules of a system, is now being used for in vivo detection of tumors (Qian et al., 2008). In another example, understanding the microbiological processes of carbonate precipitation, which is one of the main topics in the search for life on Mars, may help prevent the deterioration of buildings (Gadd, 2010).

Communication technology is also being developed on these missions, due to the need to transport large amounts of data on a small bandwidth network that is subjected to a variety of environmental interferences. Error checking and redundancy are critical to such missions, requiring special algorithms and data management schemes (Akyildiz et al., 2003) that may have spin-off applications. The SETI@home project formed the world’s largest virtual supercomputer, with over 5 million volunteers in 2011 (Anderson et al., 2002; Berkeley SETI, n.d.). This technology was a precursor for grid, cloud, and massively parallelized computing. These computing techniques are now heavily used both for research and in the private sector due to their capacity to harness the power of the multiple processors that can now be placed on a single chip. SETI@home also led to the Berkeley Open Infrastructure for Network Computing (BOINC) project (Anderson, 2004; Giorgino et al., 2010), which is an open-source software platform for computing with the use of volunteered resources. The BOINC project is currently being used in a variety of fields, including cosmology, climate prediction, drug design (designing molecules with specific objectives), and the study of various illnesses.

Astrobiology is necessarily an interdisciplinary endeavor, and it is also an international one (see Chapter 1). The requirement for astrobiologists to collaborate with others from different disciplines and other nations is stronger than in many single-discipline areas of research, precisely because astrobiology is multidisciplinary in its content. Such interdisciplinary research may help make the global research community more efficiently interconnected and promote the establishment of virtual collaboration techniques and the expansion of international scientific agreements.

Astrobiology is also providing an interdisciplinary and interconnected framework where the complexity of the Universe can be explored as a whole. As it deals with a subject that many find exciting (is there life outside Earth?) and requires study of all the different branches of science, it may help to make other scientific subjects more attractive to young people and of interest to the general public (Arino de la Rubia et al., 2009; also see Chapter 9).

8.2. What issues are raised by the possibility of contact with any life that may exist beyond Earth?

8.2.1. What strategies are in place to protect terrestrial life from potential contamination by microbial life from beyond Earth, and vice versa? Planetary protection is a principle based on the United Nations Space Treaty of 1967, and its purpose is to prevent the introduction of microbes from Earth to another celestial body and prevent the possible contamination of Earth by any life that may exist beyond our planet. Planetary protection measures are designed to prevent contamination whether this occurs intentionally or unintentionally (Horneck et al., 2007; NASA Office of Planetary Protection, 2010). Planetary protection measures have changed over time as our understanding of planetary protection issues grows. Based on the UN treaty, a concept of contamination control has been elaborated by the Committee on Space Research (COSPAR), which is linked to the United Nations Committee on Peaceful Uses of Outer Space (Debus, 2004).

The current COSPAR planetary protection guidelines group missions into five different categories (Table 9) according to the scientific interest and mission goals for the different targets in our solar system (Rummel et al., 2002; COSPAR, 2011).

To minimize the risk that spacecraft from Earth will contaminate other planetary bodies, spacecraft are normally assembled in clean rooms, and the number of microorganisms on the spacecraft (the bioburden) is monitored. The procedures to reduce the bioburden of a spacecraft include (i) surface wiping with sterilizing agents, (ii) gamma radiation sterilization, (iii) dry heat sterilization, (iv) hydrogen peroxide vapor/gas plasma sterilization (Horneck et al., 2007). The appropriate use of these procedures is determined by the physical characteristics of the spacecraft devices.

After the bioburden reduction procedures, several measurement methods may be employed. To detect microorganisms, techniques based on PCR, Live/Dead fluorescent analysis, ATP bioluminescence assay or assay for lipopolysaccharides are employed. Cultivation may also be used.

<table>
<thead>
<tr>
<th>Category</th>
<th>Type of mission</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Target body is not of direct interest for understanding the process of chemical evolution</td>
<td>NEAR-Shoemaker</td>
</tr>
<tr>
<td>II</td>
<td>Significant interest relative to chemical evolution but remote chance that contamination by spacecraft could jeopardize future exploration</td>
<td>Venus Express</td>
</tr>
<tr>
<td>III</td>
<td>Flybys and orbiter missions in which potential contamination with terrestrial life-forms could jeopardize future biological experiments</td>
<td>Mars orbiters</td>
</tr>
<tr>
<td>IV</td>
<td>Lander and probe missions in which potential contamination with terrestrial life-forms could jeopardize future biological experiments</td>
<td>Mars rovers</td>
</tr>
<tr>
<td>V</td>
<td>Earth return missions</td>
<td>Return elements of Genesis and Stardust missions</td>
</tr>
</tbody>
</table>

Table 9. COSPAR Planetary Protection Categories According to the Type of Missions to Solar System Bodies (Horneck et al., 2007; Rummel, 2007)
although many environmental microorganisms have proved impossible to culture to date. Each of these assays has limitations (Horneck et al., 2007).

There is a general consensus among scientists regarding the standard for the level of contamination considered safe on spacecraft that leave Earth to land on other planetary bodies (Rummel et al., 2002). These levels of contamination vary from less than 30 bacterial spores per lander to undetermined levels in the case of missions of category I (Horneck et al., 2007).

Missions of category V also require the prevention of back contamination and recontamination. Back contamination is the possible contamination of Earth by any life that may exist elsewhere, and recontamination is the contamination of samples retrieved from another planetary body by Earth microorganisms. The prevention of back contamination is achieved through strict quarantine procedures (Rummel, 2000; Mancinelli, 2003; Debus, 2006; Horneck et al., 2007).

Space agencies around the world are requested to follow the COSPAR guidelines, which have a major impact on their mission design, spacecraft hardware, and payload (Debus, 2006). They also necessitate reviews, detailed documentation, and management structures associated with a proposed mission (Rummel and Billings, 2004).

8.2.2. What broader sociological, philosophical, and ethical questions are raised by astrobiology and the possibility of detection of life beyond Earth? The implications of any type of contact with alien life have been extensively debated in the literature and within the scientific community. International meetings have been recently organized to address the relationship between astrobiology and topics such as history, philosophy, sociology, ethics, and religion (Dominik and Zarnecki, 2011; Dunér, 2012; Rospars et al., 2013). Within peer-reviewed scientific literature, recent reviews on these topics include a special edition of the Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences (2011, Vol. 369, Issue 1936) and the work of Chyba and Hand (2005). Recent books include the works of Vakoch (2013a, 2013b) and Jakosky (2006).


At the present time we know of only one planet that possesses life: Earth. The discovery of even a single additional data point could have significant implications for the way we understand life, its origins, and ourselves (for more, see Jakosky, 2000).

In 1961, Frank Drake created the Drake equation (http://www.seti.org/drakeequation) to conceptualize the probability that we would encounter another intelligent race (Drake and Sobel, 1992). This equation includes a “longevity” factor that accounts for the fact that intelligent life may exist but without the capacity to travel or communicate beyond its planet (Shostak, 2011). The same longevity factor also applies to the human race, as we have not had these capabilities for most of our own history. It is also the case that the risks imposed by the advancement of technology on our civilization as well as non-anthropogenic risks, such as global pandemics or meteorite impacts, may prevent us from making such a discovery in the future (Bostrom, 2002).

Much of the discussion in the media related to extraterrestrial life focuses on issues raised by the prospect of contact with an extraterrestrial civilization, also referred to as ETI (Crowe, 1997; Baum, 2010; Baum et al., 2011). However, the majority of astrobiological research is focused on the possible existence and detection of microbial life, which is the simplest, most robust, and most abundant form of life on our own planet (Whitman et al., 1998) and therefore is considered to be the most likely form of life to exist on other planets and moons.

The response of humankind to the detection of, or contact with, life beyond Earth would depend on the nature of the extraterrestrial life itself. The current search for extraterrestrial life actually encompasses two distinct endeavors: (i) the search for extraterrestrial intelligence (SETI) within our galaxy and (ii) the search for evidence of the origin and evolution of non-intelligent life within and beyond our solar system (Randolph et al., 1997). Each one has its own set of implications with several possible consequences. Astrobiologists may raise philosophical, ethical, theological, and sociological questions in an attempt to ponder some of these consequences (Dick, 2000). Some examples are those related to the three fundamental questions of astrobiology: (i) origin and evolution of life, (ii) life in the Universe, and (iii) future of life on Earth.

There is a common hope that discovering another form of life and analyzing its biochemistry and (possibly) genetics would allow us to develop a broader theory of life (Randolph et al., 1997). The existence of a second form of life would also strongly support the suggestion that life is commonplace throughout the Galaxy where Earth-like planets are present. However, these considerations raise an important question: If we find another form of life, do we have an obligation to preserve it (Persson, 2012)?

The question of whether we will ever detect life beyond Earth is separate from the question of whether such life exists, as we are unlikely to be able to investigate the majority of candidate planetary bodies in our galaxy, let alone those that are farther afield. There is also the question of whether we might miss signs of life outside Earth because we are not looking at the right time. From the study of Earth and our own solar system, we know that planetary bodies change over geological timescales. If this change over time affects habitability, it presents additional barriers to our ability to understand the distribution of life in the Universe.

A final example of the wider implications of astrobiological research is the question of the definition of life itself (see Chapter 2). This active area of debate among astrobiologists is highly relevant to ongoing research into synthetic biology and the possibility of creating an artificial life-form within a laboratory (Service, 2013), as well as the question of whether (and if so, when) computers may reach the point of being considered a living “artificial intelligence.”

Astrobiological research is, therefore, not simply investigation of the possibility of life beyond Earth but includes extensive and ongoing debate about issues that help us have a glance at different possibilities for our own future (Sagan, 1980).
Chapter 9. Resources

9.1. Organizations and websites

There are a number of organizations dedicated to the advancement of the interdisciplinary and international endeavor astrobiology has become, which include the NASA Astrobiology Program, the European Astrobiology Network Association (EANA), and other entities dedicated to the advancement of specific causes within the astrobiology community. These and other professional organizations host annual meetings and other gatherings for the communication of research, facilitation of collaborative efforts, discussion of educational activities, and enrichment of the lives of the public. More information on these organizations, including their websites, can be found in a noncomprehensive, yet extensive, list that is maintained by the NASA Astrobiology Program and located online at nai.nasa.gov/international-partners.

9.1.1. Professional organizations and journals. Astrobiology is a global endeavor with professional societies, government funding sources, community-based organizations, professional academic journals, and content-specific media outlets. There is one professional, worldwide international astrobiology society: the International Society for the Study of the Origin of Life—the International Astrobiology Society (ISSOL; http://issol.org). NASA’s Astrobiology Science Conference (AbSciCon) occurs every 2–3 years. Commission 51 of the International Astronomical Union periodically organizes the “Bioastronomy” conference and has recently partnered with ISSOL to co-organize the “Origins” conference; both of these meetings have a preponderance of astrobiology content. The EANA conference takes place every year. Many countries have astrobiology organizations, and a list can be found at nai.nasa.gov/international-partners. The independently organized astrobiology graduate conferences, AbGradCon in the United States and AbGradE in Europe, led to an informal global network of early career astrobiologists. Astrobiology-specific journals include Astrobiology, the International Journal of Astrobiology, and Origins of Life and Evolution of Biospheres. Astrobiology news outlets are available to the public through organizations and websites that include NASA’s Astrobiology page (http://astrobiology.nasa.gov), Astrobio Magazine (http://www.astrobio.net), Astrobiology Web (http://astrobio.com), the SETI Institute (http://www.seti.org), New Scientist’s astrobiology page (http://www.newscientist.com/topic/astrobiology), and the sub-Reddit community for astrobiology (http://www.reddit.com/r/astrobiology).

Finally, there is a NASA Astrobiology Strategy, published in 2015. The strategic plan was based on a broad community effort and serves the role that the Astrobiology Roadmap used to play (albeit in much more detail). This document is available online at https://astrobiology.nasa.gov/uploads/filer_public/01/28/01283266-e401-4dc8-8e05-3918b21ed879/nasa_astrobiology_strategy_2015_151008.pdf.

9.2. Education resources

9.2.1. Undergraduate and graduate courses. Many astrobiology courses have been studied for their potential to capture student interest and challenge student beliefs in a manner that provokes scientific reasoning (Offerdahl et al., 2004; Slater, 2006). Astrobiology has a track record for increasing the science literacy of college-age students, as students gain a better understanding of the nature of science by connecting concepts from different fields and by researching questions whose answers may not be definitive (Danly, 2004; Tang, 2005; Foster and Drew, 2009; Oliveira and Barufaldi, 2009). There are a variety of ways in which astrobiology courses can be arranged; literature on logical progressions for astrobiology curricula provides several different views on how to teach astrobiology, for example, those outlined by Sauterer (2000), in addition to those outlines provided in astrology textbooks such as Astrobiology: A Multi-Disciplinary Approach by Jonathan Lunine. For example, a text for a biology-based course would have a different progression of chapters and coverage of different topics than an astronomy-based course. Considerations for portfolio assessment of students in non-science-major courses such as astrobiology have also been explored (Offerdahl and Impey, 2012). For those interested in a career in astrobiology, it is generally advised that prospective students find a topic of specialty other than astrobiology and focus their research on addressing astrobiological questions. Consistent with this advice, most of the undergraduate and graduate astrobiology programs are “add-on” programs to degrees in other scientific disciplines. The NASA Astrobiology Program maintains a list of such programs, as well as a list of education and outreach resources (astrobiology.nasa.gov/education). Some European programs can be found on the EANA website (http://www.eana.net.eu).

9.2.2. Formal and informal education and outreach resources. There are many curricula available for elementary/primary and middle school grades, especially on the Internet and through science agency-sponsored websites. At the high school level, two main curricula are currently available in the United States. The first is Voyages through Time, a “year-long, integrated science curriculum for ninth or tenth grade, based on the theme of evolution and delivered on CD-ROM” (SETI Institute, 1999). The other curriculum, Astrobiology: An Integrated Science Approach, is a year-long, inquiry-based curriculum for middle school and early high school students (TERC Astrobiology, 2005). Both of these curricula, which have been thoroughly tested and published for use in precollege classrooms, are available for purchase. The NASA Wavelength directory (http://nasawavelength.org) lists space science education resources for a range of different audiences.

Examples of how to integrate astrobiology into the precollege setting are available in papers such as those by Brennan (2004), DeVore et al. (2004), Nassif and Zeller (2006), Scalice and Wilmoth (2004), and James and Van Dover (2001). There are also documented teacher professional development projects such as the Astrobiology Summer Science Experience for Teachers (ASSET) (Harman and DeVore, 2009) and an online astrobiology course for teachers (Prather and Slater, 2002). Ideas on how to integrate astrochemistry into the precollege classroom are provided by Pagni (2007). Preliminary research into astrobiology as a means to raise the science literacy of students is outlined in the work of Oliver and Fergusson (2007), while an exploration of the relationship between the nature of
Acknowledgments

This document was the product of a community of early-career researchers. The author order is such that the two co-lead editors are listed first, followed by the individual chapter editors, followed by the authors. Within each of those groups, the ordering is alphabetical and does not represent the significance or scope of individual contributions. The entire Primer team would like to thank Sherry Cady, Monika Kress, and Lindsay Hays for their insights on the document. We also would like to thank the peer reviewers that suggested critical changes to each of the individual chapters. Finally, as a community document written entirely by early-career researchers, we would like to thank the advisors that supported this work—both with resources and encouragement to participate in this community service.

Individual author acknowledgements follow. S.D.D.G. would like to acknowledge support from the NASA Astrobiology Institute’s Virtual Planetary Laboratory Lead Team, supported by the National Aeronautics and Space Administration through the NASA Astrobiology Institute under solicitations NNH05ZDA001C and NNH12ZDA002C and Cooperative Agreement NNA08CN87A and NNA13AA93A. K.E.W. acknowledges the David and Lucille Packard Foundation. A.A.B. would like to acknowledge support from Baldrich Chile and Kingston Technology, Chile. A.C. acknowledges support from the Australian National University. E.E.S. acknowledges support from the University of Washington Astrobiology Program. W.J.B. acknowledges support from the University of Washington Astrobiology Program and the NASA Astrobiology Postdoctoral Fellowship Program. R.E.A. acknowledges support from the University of Washington Astrobiology Program and the NASA Astrobiology Postdoctoral Fellowship Program. A.D.G. acknowledges support from the University of Washington Astrobiology Program and the NASA Astrobiology Postdoctoral Fellowship Program. K.L. acknowledges support from Oak Ridge Associated Universities (NSF/NSF/IGERT Program in Geobiological Systems at Montana State University. L.N. acknowledges funding by the Helmholtz Alliance “Planetary Evolution and Life” and the Intuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office through the Planet Topers alliance. This work results within the collaboration of the COST Action TD 1308. C.R.G. acknowledges support from the Carnegie Institution for Science. B.K. acknowledges support from the NASA Astrobiology Institute through a NASA Postdoctoral Program Fellowship, the NAI Early Career Research Collaboration Fellowship and the NASA Exobiology and Evolutionary Biology grant under solicitation NNX13AI08G. S.I.W. acknowledges support from the NASA Astrobiology Institute through the NASA Postdoctoral Program. J.C.L. acknowledges support from the University of Washington Astrobiology Program. K.L.L. would like to acknowledge support from the NASA Harriet Jenkins Pre-doctoral Fellowship and the CSM Bechtel K-5 Excellence in Education Initiative. C.F. acknowledges the support of an NSF graduate research fellowship. J.B. acknowledges the support of the ARC. Y.M. acknowledges the support of the CONICET graduate research fellowship. A.D.G. acknowledges support from the NASA Astrobiology Institute through the NASA Postdoctoral Program. J.S.B. acknowledges an institutional postdoctoral fellowship from the Lamont-Doherty Earth Observatory. J.B. acknowledges the support of the ARC. Y.M. acknowledges the support of the Argentinian National Council of Scientific and Technological Research (CONICET) graduate research fellowship. A.D.G. acknowledges support from the NASA Astrobiology Institute through the NASA Postdoctoral Program. X.C.A. acknowledges the support from the Argentinian National Council of Scientific and Technological Research (CONICET).


Berkeley SETI. (n.d.) SETI@home, University of California, Berkeley, CA. Available online at http://seti.berkeley.edu


Address correspondence to:
Shawn D. Domagal-Goldman
Planetary Environments Laboratory
NASA Goddard Space Flight Center
8800 Greenbelt Road
Mail Stop 699.0
Washington, MD 20771
USA

E-mail: shawn.goldman@nasa.gov

Submitted 23 December 2015
Accepted 6 June 2016