EVOLUTION OF A HABITABLE PLANET

James F. Kasting[1] and David Catling[2]

[1]Department of Geosciences, Pennsylvania State University, University Park,
Pennsylvania 16802; email: kasting@essc.psu.edu
[2]Department of Atmospheric Sciences and Astrobiology Program, University of
Washington, Seattle, Washington 98195-1640; email: davidc@atmos.washington.edu

Key Words climate history, habitable zones, planetary volatiles, rise of
atmospheric oxygen, biosignature gases

■ **Abstract** Giant planets have now been discovered around other stars, and it is
only a matter of time until Earth-sized planets are detected. Whether any of these
planets are suitable for life depends on their volatile abundances, especially water, and
on their climates. Only planets within the liquid-water habitable zone (HZ) can support
life on their surfaces and, thus, can be analyzed remotely to determine whether they
are inhabited. Fortunately, current models predict that HZs are relatively wide around
main-sequence stars not too different from our sun. This conclusion is based on studies
of how our own planet has evolved over time. Earth's climate has remained conducive
to life for the past 3.5 billion years or more, despite a large increase in solar luminosity,
probably because of previous higher concentrations of CO_2 and/or CH_4. Both these
gases are involved in negative feedback loops that help to stabilize the climate. In
addition to these topics, we also briefly discuss the rise of atmospheric O_2 and O_3,
along with their possible significance as indicators of life on other planets.

1. INTRODUCTION

The evolution of Earth's atmosphere and climate has long been the domain of geo-
scientists and biologists. Geoscientists are interested in understanding the history
of the solid planet as well as its atmosphere and ocean. Biologists are interested
in understanding biological evolution and how that may have been influenced by
interactions with the environment. Prebiotic chemists are involved with these is-
sues as well because they want to understand the conditions under which life may
have originated.

Astronomers now have reason to take an interest in this topic because they
hope to be able to observe Earth-sized planets around other stars within the next
two decades. Only giant planets—approximately 100 of them to date—have been
detected, all by the radial-velocity technique (Butler 2003, Marcy & Butler 1998).
One planet, HD 209458, has also been photometrically monitored in transit across
its parent star (Charbonneau et al. 2000). Future space missions, and perhaps
ground-based observers as well, have the potential of finding Earth-sized planets.

NASA's Kepler mission, scheduled to launch in 2006, will look for planetary transits across approximately 50,000 stars in a patch of the Milky Way (Borucki et al. 2003). In the somewhat more distant future, NASA's proposed Terrestrial Planet Finder mission (Beichman, Woolf & Lindensmith 1999), or the European Space Agency's (ESA's) proposed Darwin mission (Léger 2000, Selsis, Despois & Parisot 2002), will attempt to observe planets directly at either visible/near-infrared (IR) or thermal-IR wavelengths and, at the same time, obtain spectra of their atmospheres. With luck, we will learn not only whether other habitable worlds exist but also whether any of them are actually inhabited. It therefore behooves us to consider what it takes to create and maintain a habitable world and to consider how one might look for evidence of life on such a planet.

We should note that not everyone agrees as to what constitutes a habitable world. We concern ourselves here with the traditional definition (Dole 1964; Hart 1978, 1979): A habitable planet is one on which liquid water is stable at the surface. This definition presupposes that, like life on Earth, alien life will be carbon-based and will require liquid water. Some scientists (e.g., Hoyle 1957) have speculated that life may be a more general phenomenon and that carbon and water are not needed. This idea is indeed conceivable, but it lends itself better to science fiction than to concrete development. Many more researchers would argue that life is possible on planets (or moons) with subsurface liquid water. Mars and Europa in our own solar system are two obvious candidates, and NASA has missions in the planning stages to find out if they are inhabited. However, the search for life on such bodies can only be carried out in situ because any life that may be present does not modify the planet's atmosphere in a detectable way. In contrast, investigators can search spectroscopically for biogenic gases in the atmosphere of a planet that is habitable in the traditional sense. Hence, this traditional definition of habitability is useful to researchers looking for life on planets around other stars.

2. FORMATION OF THE ATMOSPHERE AND OCEANS

We begin by briefly discussing the formation of Earth's atmosphere and oceans and, in passing, the formation of the Earth/Moon system. Entire books have been devoted to this latter subject (Canup & Righter 2000, Hartmann, Phillips & Taylor 1986, Newsom & Jones 1990), so we do no more than scratch the surface of this topic. Our goal in this section is to establish boundary conditions for the subsequent evolution of the atmosphere and for the origin of life.

2.1. Formation of Earth and the Moon

Radiometric ($^{207}Pb/^{206}Pb$) age dating shows that Earth is approximately the same age as primitive chondritic meteorites, 4.56 ± 0.01 Ga. (We use geologists' preferred time notation as it is less cumbersome than other units. Ga stands for giga-annum or billions of years ago.) Theoretical models of planetary accretion (Canup

& Agnor 2000, Chambers & Wetherill 1998, Petit, Morbidelli & Valsecchi 1999, Safronov 1969, Wetherill 1985) suggest that the planets in the inner solar system formed on a timescale of 10^7–10^8 years. Hence, most of the action was over by ~4.5 Ga. During this initial period, growing Earth accreted numerous bodies that were the size of the Moon or larger (Wetherill 1985). Given that the Moon is approximately one eightieth of Earth's mass, these impacts were exceedingly violent: They would have vaporized any ocean that was present and melted part or all of Earth's surface (Zahnle & Sleep 1997). The largest collisions, such as the Mars-sized impact that is thought to have formed the Moon (Cameron 2000, Hartmann et al. 1986), may have driven off the atmosphere and ocean entirely—although no one has demonstrated this—and created a magma ocean with a depth of ~1000 km (Righter & Drake 1997). The oldest Moon rocks are dated at 4.44 Ga, so this singular event occurred very early in Earth's history. Radiometric dating using isotopes of Lu and Hf suggests that the Moon-forming impact occurred within 30 million years of solar-system formation (Halliday, Lee & Jacobsen 2000, Halliday et al. 2001). Moon rocks also indicate that large, 100 km diameter bodies continued to pelt both Earth and Moon until approximately 3.8 Ga (Sleep et al. 1989). This bombardment took place throughout the inner solar system, as evidenced by the large number of craters on Mercury and Mars. This turbulent first 700 million years of Earth's history is referred to as the heavy-bombardment period.

The modern concept of the formation of Earth by large impacts stands in stark contrast to earlier theories, in which Earth was thought to have formed relatively slowly by the accumulation of many small bodies. The modern theory has several implications that are important for planetary habitability. First, the occurrence of large impacts should have ensured that Earth's iron core formed at the same time as did the planet (Stevenson 1983). Indeed, hafnium-isotope evidence suggests that Earth's core formed less than 30 million years after the formation of the solar system (Yin et al. 2002). In some early models of atmospheric evolution (e.g., Holland 1962), core formation was delayed for up to half a billion years until Earth's interior was sufficiently heated by radioactive decay to allow melting to occur. In Holland's model, metallic iron remained present in the upper mantle throughout this time, implying that volcanic gases would have been much more reduced than at present. Indeed, the upper mantle probably was highly reduced in the immediate aftermath of the Moon-forming impact, as evidenced by the distribution of moderately siderophile (iron-loving) elements in mantle-derived rocks (Righter & Drake 1997, 1999). However, the mantle redox state evolved rapidly to a more oxidized level as a consequence of loss of hydrogen to space and subduction of oxidants such as ferric iron (Fe^{+3}), water, and carbonates (Catling et al. 2001, Kasting, Eggler & Raeburn, Lecuyer & Ricard 1999). For the 4 billion years for which a geological record exists, no evidence for a pronounced change in mantle redox state exists (Canil 1997, 2002; Delano 2001).

This does not rule out the possibility that small changes in mantle redox state may have occurred during the past 4 billion years. It seems probable that oxidants

continued to be exported from the crust to the mantle during the entire period (4.5–2.3 Ga) when the atmosphere remained reduced and when the rate of hydrogen escape to space was high. [There is some question about whether it was the oxidation of the mantle or the crust that mattered most in the redox transition of Earth's atmosphere. Catling et al. (2001) suggested that the accumulation of oxidants in the crust was the important factor. Everyone agrees that oxidants must have accumulated somewhere as hydrogen was lost. However, the relative degree to which oxidation of the crust or mantle affected the redox composition of gases emanating from the solid Earth is still the subject of debate.] Oxidants subducted into the mantle could have oxidized ferrous iron (Fe^{+2}) to ferric iron (Fe^{+3}) or they could have converted elemental carbon to CO_2. An increase in the mantle Fe^{+3}/Fe^{+2} ratio would have caused an increase in the mantle redox state and a corresponding increase in the oxidation state of volcanic gases. This could have have eventually helped trigger the rise of atmospheric O_2 later in Earth history, around 2.3 Ga (Holland 2002). Large amounts of elemental carbon could have buffered this redox change, however. More work needs to be done to determine what, if any, mantle redox change has occurred since 4.0 Ga.

A second aspect of Earth's violent formation that continues to affect its habitability is the formation of the Moon by a large impact. That such a large collision occurred is not particularly surprising; such events are predicted by the accretion models mentioned above. However, the fact that it occurred at just the right relative velocity and angle so as to produce a large moon is considered to be somewhat unusual. Earth-like planets in other planetary systems might therefore not possess a similar-sized moon. This could affect their subsequent climate evolution because our moon plays an important role in stabilizing Earth's obliquity. Without the Moon, and with its present, 24 h rotation period, Earth's obliquity would vary chaotically from 0° to 85° on a timescale of tens of millions of years (Figure 1) (Laskar et al. 1993). This would wreak havoc on Earth's climate and might make it difficult for such a planet to support land-based life (Ward & Brownlee 2000; but see Williams, Kasting & Caldeira 1996 for factors that might help mitigate this problem).

The effects of the Moon on Earth's climate are, in reality, even more subtle. Earth's obliquity would vary regularly if its day length was less than 12 h (Laskar et al. 1993). The Moon has been slowing Earth's rotation by tidal friction ever since it formed, so the original day length was probably well under this limit. On the other hand, the glancing blow that formed the Moon was a major reason why early Earth rotated rapidly. Furthermore, the gravitational interactions that first create chaos in this system depend on correlations between Earth's precession period and secular orbital resonances between Earth and the other planets. These resonances would presumably vary among planetary systems. We conclude that each planetary system that we might eventually observe is likely to be unique with respect to the obliquity problem and, thus, that it would be inadvisable to make sweeping statements about planetary climates on the basis of what we calculate for Earth.
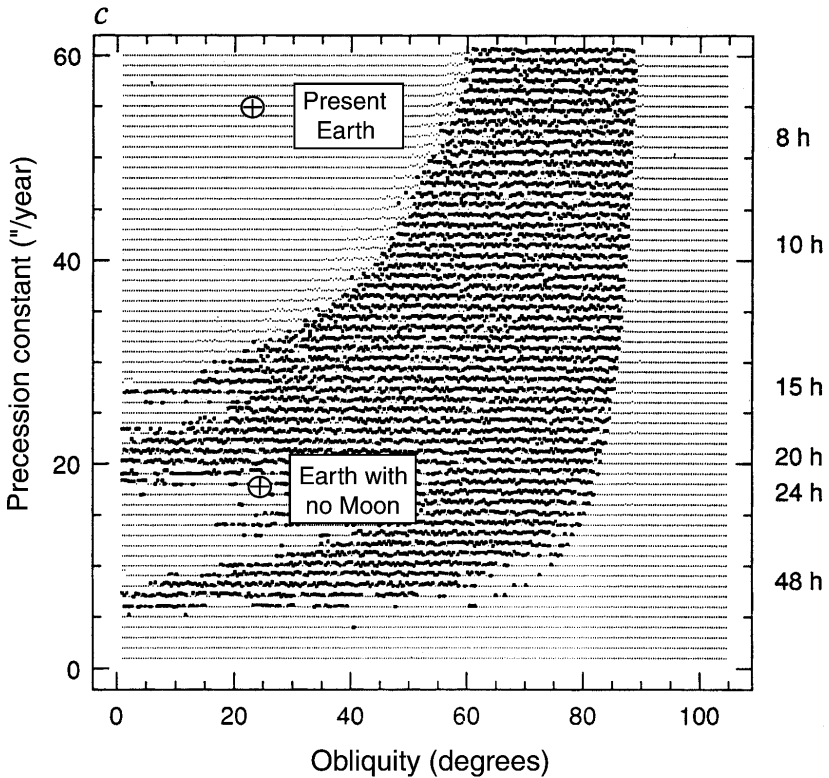
Figure 1 Regions of chaotic (*fuzzy lines*) and regular (*clear areas*) obliquity variations for Earth. The vertical axis is the precession constant, which is inversely proportional to planetary spin rate for the case of no moon. The fuzzy lines represent obliquity variations experienced during a series of 18-million-year integrations starting from different initial obliquities and precession constants. The presence of the Moon increases the torque on Earth's equatorial bulge, which in turn increases the precession constant. (Modified from Laskar & Robutel 1993.)

2.2. Delivery of Water and Other Volatiles

Modern concepts of planetary accretion also affect how we think about the formation of early Earth's atmosphere and ocean. In early models (e.g., Holland 1962, Rubey 1955), the atmosphere and ocean formed by outgassing of volatiles from Earth's interior. Rubey, a geologist, coined the term excess volatiles to describe those compounds (H_2O, CO_2, N_2, HCl, various sulfur compounds) that are present on Earth's surface in abundances that are too high to have been produced by the weathering of igneous rocks and their subsequent conversion to sedimentary rocks. In his view, these excess volatiles must have been supplied by volcanism. Rubey's

concept of excess volatiles is still a useful way to categorize surface inventories; however, we now think that many of these volatiles may have been delivered to the surface during accretion by a process termed impact degassing (Lange & Ahrens 1982; Matsui & Abe 1986a,b). At least some of the large bodies that formed Earth should have been volatile rich. Once Earth had reached approximately one third of its final mass, the acceleration caused by its own gravity would have been sufficient to ensure that the impacting bodies were partially or totally vaporized. Consequently, the atmosphere and ocean should have started to form as the planet formed. Although some volatiles must have been stored directly in Earth's interior, as evidenced by continued outgassing of primordial ^{3}He (Craig et al. 1975), most of the volatile material was probably deposited at the surface. This implies that, on average, Earth was ingassing, rather than outgassing, volatiles during much of its history. Indirect support for ingassing comes from comparing outgassing fluxes of carbon to ^{36}Ar, which are in the ratio of $\sim 8 \times 10^9$, to the ratio of carbon to ^{36}Ar in the ocean, atmosphere, and crust, which is only $\sim 1.8 \times 10^6$. This suggests that carbon has been returned to the mantle (Walker 1990).

Of critical importance to Earth's habitability is its large (1.4×10^{21} kg) ocean, which covers 70% of its surface to an average depth of 4 km. To estimate whether other Earth-like planets might possess similar amounts of water, one needs to understand where Earth's water came from. This question has been fraught with controversy. The current consensus is that the planetesimals that formed at Earth's orbital radius were largely devoid of water because the solar nebula was too hot for ice to be present and because gas-phase hydration of silicate minerals is slow (Prinn & Fegley 1989). However, not all the planetesimals that formed the bulk Earth originated from within 1 AU. The models that predict large impacts also predict substantial mixing of material between accretion zones as large planetesimals interacted gravitationally and pumped each other up to high eccentricities (Canup & Agnor 2000, Wetherill 1985). Dynamical calculations by Morbidelli et al. (2000) suggest that one or more large planetary embryos formed beyond 2.5 AU may have been swept up by the accreting Earth as a consequence of this mixing. Meteorites from the outer asteroid belt, 2.5–3.5 AU, have approximately the same average deuterium/hydrogen ratio as do Earth's oceans, so the prediction that Earth obtained much of its water from this region is consistent with isotopic constraints. Icy bodies from the Jupiter-Saturn region should also have come in early, as these objects have very short dynamical lifetimes (Morbidelli et al. 2000). Their hydrogen isotopic composition is unknown but would arguably not have been too different from that of outer-belt asteroids.

Some additional water must also have come in during the heavy-bombardment period. Asteroids parked in highly inclined orbits may have continued to arrive in the inner solar system for hundreds of millions of years (Morbidelli et al. 2001). Comets from the Uranus-Neptune region would also have been late to arrive. Several different authors (e.g., Chyba 1989, Delsemme 2001) have suggested that these comets were a major source of Earth's water. Such low-temperature comets could have had noble-gas contents similar to those found on Earth, whereas

meteorites are embarrassingly rich in xenon relative to argon and krypton (Owen et al. 1992). However, Oort Clouds comets (which formed in this region) have deuterium/hydrogen ratios roughly twice that of Earth [$(310 \pm 40) \times 10^{-6}$ versus $(149 \pm 3) \times 10^{-6}$], based on observations of Halley, Hale-Bopp, and Hyakatake (Robert 2001). Morbidelli et al. (2000) found that at most 10% of Earth's water could have come from such objects. Their model is attractive, provided one relaxes the stipulation that delivery of water should be linked directly to delivery of the noble gases.

The implications of this story for Earth-like planets orbiting other stars are unclear. On one hand, the dynamical scattering of water-rich bodies from the outer solar system is something that would presumably happen in all planetary systems. Thus, all terrestrial planets would be expected to acquire at least some fraction of Earth's water. On the other hand, the accretion of large, planetary embryos is a stochastic process. In the numerical simulations of Morbidelli et al. (2000), some terrestrial planets accreted water-rich embryos, whereas others did not. The timing of the delivery of volatiles, the size of asteroid-like volatile reservoirs, and the efficiency with which the volatiles are delivered depend sensitively on the locations, sizes, and initial orbital eccentricities of the giant planets (Chambers & Wetherill 2001). Hence, planets similar to Earth but possessing significantly less water may well exist in other planetary systems.

2.3. Effects of Large Impacts on the Atmosphere and on Life

Regardless of whether the heavy-bombardment period was responsible for the delivery of Earth's volatiles, it should have affected the composition of the early atmosphere and may have influenced the origin of life. The larger impacting bodies would have vaporized substantial amounts of water, thereby creating a steam atmosphere. Such an atmosphere may have been present more or less continuously during the main accretion period (Matsui & Abe 1986a,b), but later it would have existed only transiently following large impacts. Evidence that this actually happened is provided by neon isotopes. Gases obtained from mantle minerals and from deep wells have a solar $^{22}Ne/^{20}Ne$ ratio, whereas Earth's atmosphere has a ratio that is nearly ten times higher (Pepin 1991). This is most easily explained by rapid, hydrodynamic escape of hydrogen at a rate sufficient to drag off ^{20}Ne (Zahnle et al. 1990). A hydrogen-rich upper atmosphere is required for this to have occurred. Furthermore, the escape must have happened early in Earth's history when the solar UV flux was much higher than it is today (Walter & Berry 1991, Zahnle & Walker 1982). Both conditions would have been met in an impact-produced steam atmosphere. Escape of hydrogen during this period may also explain why Earth's mantle appears to have been oxidized since very early in its history (see Section 2.1). Although it has not been demonstrated quantitatively, enough hydrogen could have been lost during this period to raise the mantle to its current oxidation state.

The large impacts occurring during the heavy-bombardment period may also have affected the timing of life's origin. The amount of time between major impacts

increases significantly between 4.5 Ga and 3.8 Ga (Maher & Stevenson 1988). Thus, if a certain period of uninterrupted chemical evolution was required before life could begin, the process would have become increasingly likely as time went on. An alternate, and perhaps better-defined, way of accounting for the effects of impacts is to ask when the last sterilizing impact event occurred (Sleep et al. 1989). Impactors with diameters in excess of ~100 km should have had sufficient kinetic energy to vaporize the photic zone (the uppermost 100 m) of the oceans, whereas those with diameters of 440 km or greater could have vaporized the entire ocean. For reasonable assumptions about the size distribution of planetesimals, the last ocean-vaporizing event probably occurred around 4.1–4.3 Ga (Sleep et al. 1989), although one cannot rule out an event as late as 3.8 Ga (Figure 2). Photic-zone
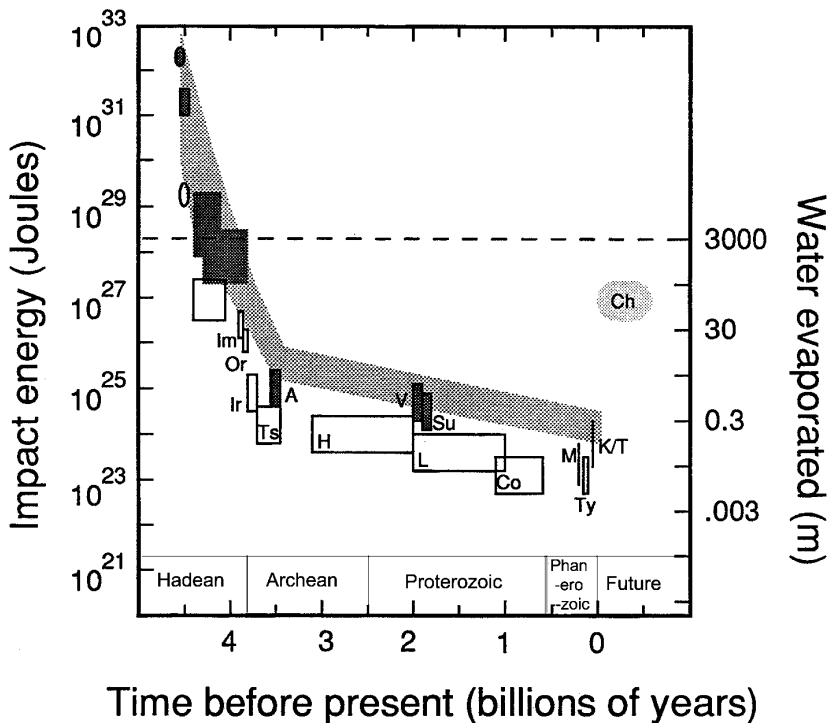

Figure 2   Energies released by impacts on the Moon (*open boxes*) and on Earth (*filled boxes*). Lunar craters are Tycho, Copernicus, Langrenus, Hausen, Tsoilkovski, Iridum, Orientale, and Imbrium. Terrestrial events are the K/T impact, Manicougan, Sudbury, Vredevort, and an impact energy corresponding to the thickness of Archean spherule beds. Ovals are self-energies of formation; the early box refers to the Moon-forming impact. The stippled region represents the range of impact energies on Earth inferred from the lunar data. The depth of ocean evaporated is also given. The dashed line corresponds to an ocean-vaporizing event. (From Sleep et al. 1989.)

evaporators are expected to have hit Earth until 3.8 Ga, as several large lunar basins, including Imbrium and Orientale, were formed at approximately this time.

Whether life originated during the heavy-bombardment period is uncertain. Claims that a diverse assemblage of bacteria existed in Australia 3.5 Ga (Schopf 1993) have been recently disputed (Brasier et al. 2002). However, microfossils from South Africa dated at 3.23–3.47 Ga consist of spheroidal carbonaceous objects, some of which are in a process of binary division similar to that of cells and are plausibly biogenic (Knoll & Barghoorn 1977). Some investigators have recently disputed the claim that isotopically fractionated organic carbon, or kerogen, dated at 3.8–3.9 Ga (Mojzsis et al. 1996) is indicative of life (Fedo & Whitehouse 2002, van Zuilen et al. 2002). However, a similar claim by Rosing (1999) may still be viable.

An early origin for life could explain one of the most intriguing features of the biological record: In evolutionary trees derived from sequencing of ribosomal RNA, most of the organisms near the root of the tree are hyperthermophiles with preferred growth temperatures in excess of 80°C (Pace, Olsen & Woese 1986). One explanation for this finding is that life originated at high temperatures, perhaps in some midocean-ridge hydrothermal-vent system (Corliss, Baross & Hoffman 1981). However, this inference is vigorously contested by some prebiotic chemists who argue that life must have originated at lower temperatures at which amino acids and other organic precursor molecules are more stable (Bada, Bigham & Miller 1994). An alternate explanation for this observation is that life originated during the heavy-bombardment period. Life could have originated at low temperatures, then colonized the midocean-ridge vent systems. A large impact may have subsequently wiped out all surface life, leaving hyperthermophilic vent organisms to recolonize the entire planet (Sleep et al. 1989). This hypothesis can explain the phylogenetic data without requiring a high-temperature origin of life.

3. LONG-TERM CLIMATE EVOLUTION

What would Earth's climate have been like in the distant past? To address this question, we first briefly review the fundamental equations that govern a planet's mean surface temperature. We apply them to modern Earth and then consider their implications for an early Earth warmed by a dimmer young sun.

3.1. Planetary Energy Balance and the Greenhouse Effect

Earth is warmed by absorption of visible and near-IR radiation from the Sun and is cooled by emission of thermal IR radiation. If we treat Earth as a blackbody with effective temperature T_e, the condition of planetary energy balance can be readily shown as

$$\sigma T_e^4 = \frac{S}{4}(1 - A). \tag{1}$$

Here, σ ($= 5.67 \times 10^{-8}$ W/m^2/K^4) is the Stefan-Boltzmann constant, S ($=$ 1370 W/m^2) is the solar flux at Earth's orbit, and A ($\cong 0.3$) is the planetary albedo

or reflectivity. Solving for the effective temperature yields $T_e = 255$ K. In reality, of course, Earth is not a blackbody. Instead, it has an atmosphere that warms the surface by way of the greenhouse effect. Infrared radiation emitted by the planet's surface is absorbed and re-emitted by IR-active gases within the atmosphere. The extra downward IR radiation helps to warm the surface. Thus, the global average surface temperature T_s is $\sim$288 K. The difference between T_e and T_s is the magnitude of the greenhouse effect:

$$\Delta T_g \equiv T_s - T_e = 33 \text{ K}. \tag{2}$$

In Earth's atmosphere today, the two most important greenhouse gases are CO_2 and H_2O. H_2O is responsible for approximately two thirds of this warming; CO_2 accounts for most of the remaining one third of the greenhouse effect. Lesser contributions, on the order of two to three degrees total, come from CH_4, N_2O, O_3, and various anthropogenic chlorofluorocarbons (CFCs).

The most difficult factor to deal with in Equations 1 and 2 is the planetary albedo A. This is because most of it (approximately 0.25 of the 0.3 total) is caused by clouds. Clouds can be observed and parameterized in Earth's present atmosphere, but their properties are difficult to predict for atmospheres other than our own. For this reason, climate calculations for early Earth or for other Earth-like planets are subject to considerable uncertainty. However, this does not mean that such calculations are meaningless. Water (or ice) clouds on a planet with a mean surface temperature and fractional ocean cover similar to that of present Earth would presumably have much the same average characteristics as those we observe on Earth. If the goal is to determine what is required to create a climate similar to that of today, it is reasonable to assume no change in cloud properties. For model planets that are either much hotter or much colder than present Earth, however, the neglect of cloud feedback may lead to serious errors.

3.2. Climate Feedbacks and Feedback Loops

Water vapor behaves differently than CO_2 does in Earth's atmosphere because it is near its condensation temperature. Hence, it acts as a feedback, rather than a forcing, on the climate system. (CO_2 can also act as a feedback mechanism but on much longer timescales.) The water-vapor feedback is nearly instantaneous. If the climate cools, the saturation vapor pressure drops, and the atmospheric water-vapor concentration decreases proportionately. Less water vapor results in a smaller greenhouse effect, which results in further cooling. Just the opposite happens if the climate warms: Atmospheric H_2O increases, thereby increasing the greenhouse effect and amplifying the initial warming.

A second important climate feedback is the snow/ice-albedo feedback. An increase in surface temperature causes a decrease in snow and ice cover, thereby decreasing the planetary albedo, which causes an increase in surface temperature. This amplifies the initial warming, so the feedback loop is again positive. Researchers believe that the snow/ice-albedo feedback loop has played a major role in the advances and retreats of the polar ice sheets throughout the past 2 million years.

The climate system must also contain negative feedbacks or it would be unstable. The most basic negative feedback is the interaction between surface temperature and the outgoing IR flux (F_{IR}). As T_s increases, F_{IR} increases. However, Earth cools itself by emitting IR radiation; thus, as F_{IR} increases, T_s decreases. This creates a negative feedback loop so fundamental that it is often overlooked. This feedback loop is the reason Earth's climate is stable on short timescales. On long timescales, the factors that affect climate (e.g., solar luminosity) can change; thus, the IR feedback loop by itself no longer ensures stability. The next section describes a negative feedback that is critical to Earth's long-term habitability and may well be important for Earth-like planets around other stars.

3.3. The Carbonate-Silicate Cycle and the CO₂-Climate Feedback

What controls atmospheric CO_2 concentrations over long timescales? CO_2 is controlled by the carbon cycle, but this cycle has several different parts. The one with which most people are familiar is the organic carbon cycle, in which plants (and many microbes) convert CO_2 and H_2O into organic matter and O_2 by photosynthesis: $CO_2 + H_2O \rightarrow CH_2O + O_2$. Photosynthesis is balanced by respiration and decay, which reverse the process and recreate CO_2 and H_2O.

However, there are two reasons why the organic carbon cycle cannot be the primary control on CO_2 levels over long timescales. First, the living biosphere is not a large carbon reservoir—it only contains approximately as much carbon as does the atmosphere. Thus, substantial imbalances in the cycle cannot be maintained. Second, the small imbalance that does occur when organic carbon is buried in sediments cannot be maintained for very long because it is controlled by a negative feedback loop involving atmospheric O_2. A single geochemical cycle does not have enough degrees of freedom to regulate simultaneously both atmospheric O_2 and climate.

The most important part of the carbon cycle in terms of long-term climate is the inorganic carbon cycle, sometimes called the carbonate-silicate cycle As shown in Figure 3, CO_2 dissolves in rainwater to form carbonic acid (H_2CO_3), which is a weak acid, but when it acts over long timescales, it is strong enough to dissolve silicate rocks. For illustrative purposes, we use the simplest silicate mineral, wollastonite ($CaSiO_3$), to represent all silicate rocks. The products of silicate weathering, including calcium (Ca^{2+}) and bicarbonate (HCO_3^-) ions and dissolved silica (SiO_2), are transported by streams and rivers to the ocean. There, organisms, such as foraminifera, use the products to make shells of calcium carbonate ($CaCO_3$). Limestone is the commonly preserved form of calcium carbonate. Other organisms such as diatoms and radiolarians make shells out of silica. When these organisms die, they fall into the deep ocean. Most of the shells redissolve, but a fraction of them survive and are buried in sediments on the seafloor. The combination of silicate weathering plus carbonate precipitation can be represented chemically by $CO_2 + CaSiO_3 \rightarrow CaCO_3 + SiO_2$.
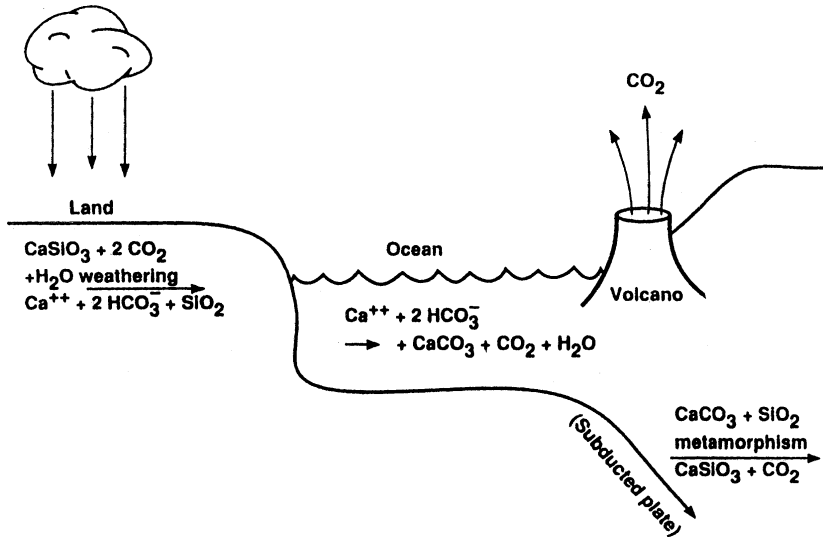
Figure 3 Diagram illustrating the modern carbonate-silicate cycle, also referred to as the inorganic carbon cycle.

If silicate weathering and carbonate precipition were the only reactions occuring, all of Earth's CO_2 would eventually wind up in the carbonate rock reservoir and the planet would become uninhabitable. Fortunately, another part to the cycle exists. As we know from the theory of plate tectonics, the seafloor is not static. Rather, it is continuously created at the mid-ocean ridges, and it is subducted at certain plate boundaries when the denser oceanic plate dives beneath the less dense continental plate. When this happens, the overlying carbonate sediments are carried down to depths where the temperatures and pressures are much greater. Under these conditions, carbonate minerals recombine with SiO_2 (which by this time is the mineral quartz) to reform silicate minerals, releasing CO_2 in the process. This reaction is termed carbonate metamorphism. The CO_2 released from carbonate metamorphism makes its way back to the surface and re-enters the atmosphere by way of volcanism, thereby completing the carbonate-silicate cycle. This cycle replenishes all the CO_2 in the combined atmosphere-ocean system on a timescale of approximately half a million years.

The carbonate-silicate cycle contains a negative feedback that stems from the dependence of the silicate weathering rate on surface temperature, T_s (Berner, Lasaga & Garrels 1983, Walker, Hays & Kasting 1981). Weathering rates increase both because of the direct effect of temperature on chemical reaction rates and because evaporation (and, hence, precipitation) rates increase as T_s increases. As silicate weathering is the loss process for atmospheric CO_2, CO_2 concentrations should tend to fall as T_s rises and CO_2 should increase as T_s falls. The response time of this feedback loop is that of the carbonate-silicate cycle—hundreds of

thousands to millions of years. It is thus too slow to counteract human-induced global warming, but fast enough to have a dominating effect on the billion-year timescale of planetary evolution.

3.4. The Faint Young Sun Problem

Beginning in the 1950s with the work of astrophysicsts such as Hoyle (1958) and Schwarzschild (1958), astronomers developed theoretical models for how stars evolve over time. One prediction of virtually all such models is that the Sun, like other stars, gets brighter as it ages. The Sun produces energy by fusing four H nuclei into one helium(He)-4 nucleus deep within its core. As it does so, the density of the core increases. The Sun is a self-gravitating sphere; thus, the increase in core density causes it to shrink slightly and heat up. Higher core temperatures cause the fusion reactions to proceed faster, so the Sun produces increasingly more energy. Its luminosity must therefore increase with time to maintain energy balance. Standard solar models predict that the Sun was approximately 30% less bright when it formed at 4.6 Ga and that its luminosity has increased more or less linearly with time (Gough 1981) (Figure 4).
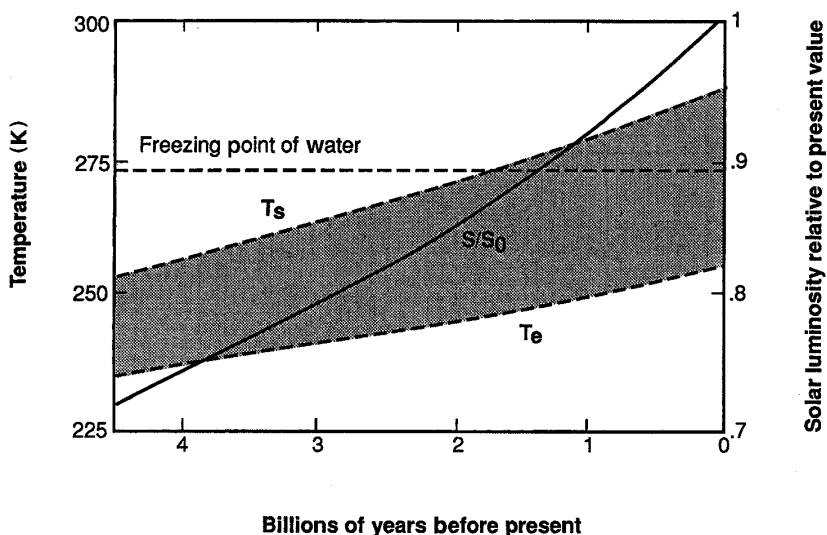

Billions of years before present

Figure 4 Diagram illustrating the faint young Sun problem. The solid curve represents solar luminosity relative to the present value, as predicted by the computer model of Gough (1981). The lower dashed curve is Earth's effective radiating temperature, T_e, computed by a one-dimensional, radiative-convective climate model. The upper dashed curve represents the calculated mean global surface temperature, T_s. A fixed CO_2 mixing ratio of 300 ppmv and a fixed relative humidity were assumed in the calculation. (From Kasting, Toon & Pollack 1988).

Hoyle (1958) made this prediction back in the late 1950s; however, it was not until 14 years later that Sagan & Mullen pointed out the implications for planetary climates (1972). If one reduces the value of S by 30% in (1), holding A and ΔT_g constant for simplicity, one finds that T_e drops to 233 K and $T_s = 266$ K, well below the freezing point of water. If the calculation is repeated with a climate model that includes the positive feedback loop involving water vapor, the problem becomes even more severe. The dashed curves in Figure 4 show T_e and T_s calculated using a one-dimensional, radiative-convective climate model, assuming constant CO_2 concentrations and fixed relative humidity (Kasting, Toon & Pollack 1988). The results are remarkably similar to those predicted earlier by Sagan & Mullen: T_s drops below the freezing point of water prior to ~2 Ga. Combined with the snow/ice-albedo feedback loop, this temperature drop would almost certainly lead to a globally glaciated Earth. However, geologic evidence tells us that liquid water and life were both present as far back as 3.5 Ga and maybe longer. The oldest zircons, zirconium silicate minerals that must have formed in liquid water, are dated at more than 4.3 Ga and may indicate the presence of an ocean at that time (Catling & Kasting 2002, Mojzsis, Harrison & Pidgeon 2001, Wilde et al. 2001).

How can the faint young Sun problem be solved? A large decrease in cloudiness would do it (Rossow et al. 1982), but this seems unlikely for reasons mentioned in Section 3.1. Instead, the answer probably lies in increased concentrations of greenhouse gases. Both CO_2 and CH_4 are plausible candidates. One can think of good reasons why both these gases should have been more abundant in the distant past.

For CO_2, the argument involves the carbonate-silicate cycle discussed in the previous section. As an extreme example, suppose that the oceans were completely frozen. (This is not a completely hypothetical situation, as this may have actually happened during "Snowball Earth" episodes.) In this scenario, the hydrological cycle would have shut down almost completely. Some H_2O would still cycle by sublimation and snowfall, but liquid water would be completely absent. Weathering of rocks requires liquid water to proceed at an appreciable rate, so the rate of silicate weathering would drop to near zero. Volcanic activity, however, would continue unabated. It appears in Figure 3 as if the volcanic CO_2 source might disappear, too, but that is not the case. The mean residence time of seafloor is 60 million years, and besides, other types of volcanism release CO_2 as well (e.g., midocean-ridge and hotspot volcanism). Thus, volcanic CO_2 would accumulate in the atmosphere until, eventually, the greenhouse effect became strong enough to melt the ice. Approximately 0.3 bars of CO_2 would be required to melt the ice of a completely frozen Earth (Caldeira & Kasting 1992). At present rates of volcanism, this much CO_2 would have taken ~10 million years to accumulate.

Although there may indeed have been a few times in Earth's history when such an extreme scenario actually occurred (see below), most of the time Earth has remained in the liquid-water regime. As long as the silicate weathering rate slows down as the climate grows colder, the negative feedback loop described in Section 3.3 should help stabilize the climate. This same stabilizing feedback would

presumably operate on other planets as well, provided that they had sufficient stocks of carbon and water and enough internal heat to maintain active plate-tectonic recycling of CO_2.

3.5. Failure of Climate Stabilization on Mars and Venus

One way of evaluating the importance of the CO_2-climate feedback loop is by examining what happened to our neighboring planets, Venus and Mars. Neither planet has a surface that is habitable, according to our previous definition. Venus has a mean surface temperature of approximately $460°C$, well above the critical point for water, whereas Mars' mean temperature is approximately $-55°C$, which is well below freezing. Why were their climates not stabilized by the carbonate-silicate cycle?

It is not too difficult to determine what went wrong on these planets. Venus was close enough to the Sun that it lost its water by the process mentioned previously, namely, photodissociation followed by escape of H to space. The critical solar flux for water loss in a model with zero cloud feedback is 1.1 S_0, where S_0 is the current flux at Earth's orbit (Kasting, Toon & Pollack 1988). The flux at Venus' orbit today is 1.91 S_0, and the flux at Venus' orbit at 4.5 Ga was ~1.4 S_0, so Venus appears to have been susceptible to water loss throughout its history. Once the water was lost, silicate weathering could not occur, so volcanic CO_2 simply accumulated in Venus' atmosphere. The result was the hot, dry planet that we see today. Arguments that the present Venusian climate is unstable (Bullock & Grinspoon 1996) seem implausible because of the extremely slow rate of reaction between atmospheric gases and dry rock.

In the case of Mars, the problem was quite different. Mars formed farther from the Sun, so that a stronger atmospheric greenhouse effect would have been needed to warm its surface. Gaseous CO_2 and H_2O by themselves could not have kept Mars warm early in its history because both gases would have condensed out of the atmosphere owing to the solar flux prevailing at that time (Kasting 1991). Radiative warming by CO_2 ice clouds might have allowed Mars to overcome this difficulty for some time (Forget & Pierrehumbert 1997). However, Mars had another, more serious problem: It has only approximately one ninth of Earth's mass. Thus, its interior cooled more rapidly than Earth's, and widespread volcanism ceased a long time ago. Without volcanism, there was no mechanism for recycling CO_2, so most of the CO_2 that was present should have accumulated in the crust. Admittedly, this theory is flawed in that despite several spectroscopic searches, carbonate minerals have not been found on Mars. An idea that explains this observation is that, because of low Martian gravity, the early Martian atmosphere was eroded away into space by comet or asteroid impacts (Melosh & Vickery 1989). It may be that other greenhouse gases (CH_4?) were needed to keep early Mars warm. One lesson that can be drawn from Mars, however, is that planetary size matters: A geologically inactive planet is unlikely to be able to maintain a stable climate or hold onto its atmosphere for geologically long periods.

3.6. The Habitable Zone Around the Sun and Other Stars

The climate calculations described above can be used to estimate the liquid wa-
ter habitable zone (HZ) around the Sun and other main-sequence stars (Kasting,
Whitmire & Reynolds 1993). These authors also calculated the width of the contin-
uously habitable zone (CHZ), which is the region that remains habitable for some
finite period of time. Contrary to conclusions reached earlier by Hart (1978, 1979),
Kasting et al. (1993) found that HZs and CHZs around most solar-type stars were
relatively wide (Figure 5). The major difference from Hart's studies is the inclusion
in the latter model of the stabilizing feedback provided by the carbonate-silicate
cycle. When this negative feedback is included, the inner edge of the present HZ
around our sun is predicted to lie at ~0.95 AU (the water loss limit), whereas the
outer edge lies somewhere between 1.37 AU and 2.4 AU, depending on the amount
of warming provided by CO_2 ice clouds (Forget & Pierrehumbert 1997, Mischna
et al. 2000). HZ's around other main-sequence stars appear to be approximately
the same width when distances are expressed on a log scale (Figure 5). If planets
in general are geometrically spaced, as they are in our solar system, the chances of
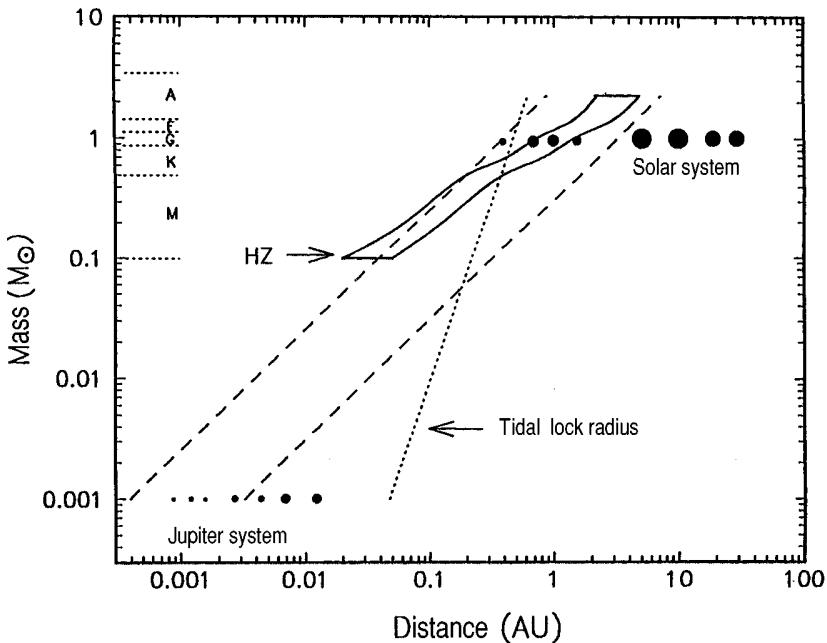

Figure 5   The zero-age main-sequence habitable zone (HZ) around different types of
stars. Runaway-greenhouse and maximum-greenhouse limits on the HZ are assumed.
The nine planets of our own solar system are shown. The dashed curve is the distance
within which a planet's rotation is predicted to become tidally locked during a period
of 4.5 billion years. (From Kasting, Whitmire & Reynolds 1993).

finding other habitable planets are relatively good. The chances diminish for bright blue stars because of their high UV fluxes and short main-sequence lifetimes. Dim red stars are also less likely to host habitable planets because their HZs are close enough to the stars that planets within them would be tidally locked. This could cause the atmospheres and oceans of dim red stars to freeze on their dark sides. M-star planets with dense atmospheres might avoid this problem by efficiently transporting heat from their daysides to their nightsides (Joshi et al. 1997).

3.7. Complications to the Faint Young Sun Problem: The Importance of CH_4

The explanation provided above for how early Earth avoided freezing is intuitively attractive. However, that does not mean that CO_2 increases and the silicate weathering feedback are the only factors that helped keep the early Earth warm. As discussed below, atmospheric O_2 concentrations are thought to have been low until $\sim$2.3 Ga. In a low-O_2 atmosphere, various reduced greenhouse gases, such as methane (CH_4) and ammonia (NH_3), could conceivably have been more abundant. Indeed, in their 1972 paper, Sagan & Mullen proposed high concentrations of NH_3 as the solution to the faint young Sun problem. This idea now seems unlikely because NH_3 photolyzes rapidly in the absence of UV screening by atmospheric O_2 and O_3 (Kasting 1982, Kuhn & Atreya 1979), even when shielding by hydrocarbon haze is taken into account (Pavlov, Kasting & Brown 2001). CH_4, though, is not subject to this problem, as it photolyzes only at wavelengths below $\sim$145 nm. Photochemical models (Pavlov, Kasting & Brown 2001, Zahnle 1986) predict that the lifetime of CH_4 in a low-O_2 atmosphere is $\sim$5,000–10,000 years, as opposed to $\sim$12 years today. Thus, the present biological methane flux of 535 Tg CH_4/year (Houghton et al. 1994) could have supported an atmospheric CH_4 mixing ratio of $\sim$3 $\times$ 10^{-4} (Pavlov, Kasting & Brown 2001), instead of the 1.6 $\times$ 10^{-6} that it supports today. (The term mixing ratio means the same as mole fraction.)

This by itself does not prove that CH_4 was abundant in the early atmosphere, of course, because there is no reason to suppose that methane was produced at the same rate then as today. Indeed, the Hadean (pre-3.8 Ga) methane flux may have been small because life may not yet have originated, or at least proliferated, by this time. (The geologic eras of life are shown in Figure 2.) However, the Archean biogenic methane flux was probably substantial. The methanogenic bacteria (or methanogens for short) that make methane today are thought to be evolutionarily ancient, based on sequencing of ribosomal RNA (Woese & Fox 1977). Furthermore, the substrates used by methanogens—H_2, CO_2, acetate, and formate—are all thought to have been abundant during the Archean. Virtually all methanogens can subsist by way of the reaction: $CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$. Thermodynamic considerations suggest that methanogens should have converted most of the available atmospheric H_2 into CH_4 by this pathway (Kasting, Pavlov & Siefert 2001, Kral et al. 1998). Abiotic H_2 mixing ratios are thought to have been of the order of 10^{-3}, so postbiotic CH_4 mixing ratios should have been roughly the same.

(Actually, total H concentrations are preserved in this process, so the predicted CH_4 mixing ratio is half that of H_2.) This estimate is consistent with the one obtained in the previous paragraph, although the argument used is entirely different.

If CH_4 was an abundant constituent of the Archean atmosphere, then the greenhouse effect could have been large even if CO_2 concentrations were relatively low. Figure 6 illustrates this point quantitatively. The solid curves in Figure 6 represent global mean surface temperatures calculated with a one-dimensional, radiative-convective climate model (Pavlov et al. 2000). These calculations were performed for a time during the Late Archean, 2.8 Ga, when the solar flux was ~80% of its present value. The dashed curves represent the freezing point of water (273 K) and an upper limit on pCO_2 derived from paleosols (Rye, Kuo & Holland 1995). Paleosols are ancient soils that have been preserved and compacted to form rock. Looking for the presence of the mineral siderite ($FeCO_3$), Rye, Kuo & Holland (1995) examined paleosols formed from 2.2–2.8 Ga. This mineral should have formed if CO_2 were as abundant in the early atmosphere as simple climate models predict; that it did not implies that pCO_2 was below the indicated level. The

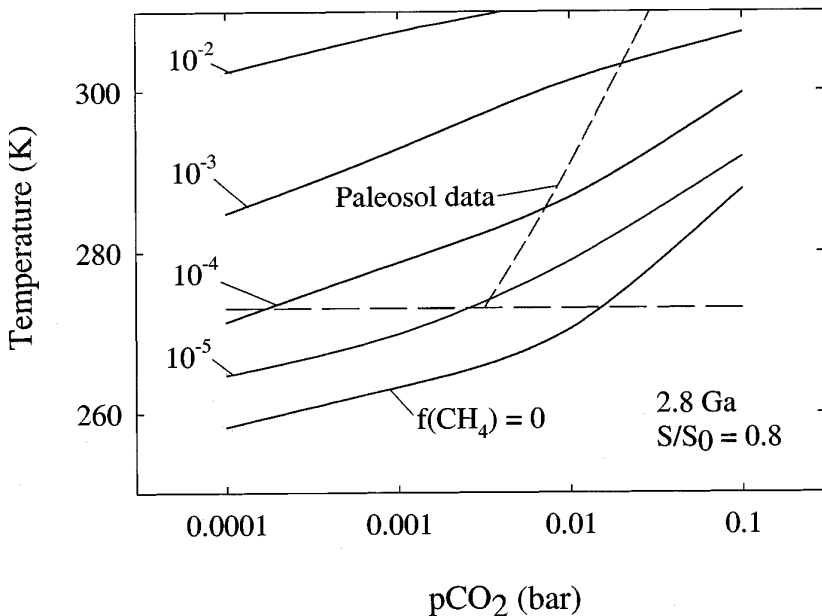**Figure 6**   Global mean surface temperature T_s as a function of atmospheric CO_2 partial pressure (pCO_2) and CH_4 volume mixing ratio. The calculations were performed for a solar luminosity equal to 80% of the present value, which is the value predicted for 2.8 Ga. The dashed curves represent the freezing point of water and the upper limit on pCO_2 derived from paleosol data (Rye, Kuo & Holland 1995). (From Pavlov et al. 2000.)

constraint takes the form of a curve, rather than a vertical line, because the reactions involved in siderite formation depend on temperature.

Figure 6 shows that in the absence of CH_4, the CO_2 partial pressure required to keep T_s above freezing was at least 0.2 bars or $\sim$20 times greater than the upper limit derived from paleosols. However, if the CH_4 mixing ratio were 10^{-4}–10^{-3}, as suggested above, then the surface could have been warm even if pCO_2 was no higher than today. Most methanogens are thermophilic (heat-loving), and those with higher optimum growth temperatures have shorter doubling times than those that grow best at lower temperatures (Cooney 1975). This creates a positive feedback loop that should have tended to keep the Archean climate warm. This positive feedback was limited, though, by an additional complication: When CH_4 becomes as or more abundant than CO_2 in a planet's atmosphere, photochemical models (Pavlov, Kasting & Brown 2001) predict that it can polymerize to form hydrocarbon haze similar to that observed today on Saturn's moon Titan. This haze, in turn, can cool the planet's surface via the antigreenhouse effect (McKay, Pollack & Courtin 1991, Pavlov, Kasting & Brown 2001). In the antigreenhouse effect, solar radiation is absorbed and reradiated high in the atmosphere without ever having a chance to make it down to the surface. If the haze had become too thick, the surface would have frozen, and conditions would have become unfavorable for the existence of methanogens. This negative feedback loop may have stabilized the Archean climate and maintained a persistent, optically thin haze layer (Pavlov et al. 2001).

3.8. End of the Methane Greenhouse: The First Snowball Earth

We end our discussion of planetary climates approximately halfway through Earth's history at the time when atmospheric O_2 levels first rose. As discussed below, this event appears to have occurred around 2.3 Ga. If methane was an important contributor to the greenhouse effect during the Late Archean, then the rise of O_2 must have been significant for the climate. Indeed, the correlation between the two was noted more than 30 years ago by the Canadian geologist Roscoe (1969, 1973). He observed rocks in the Huronian sequence just north of Lake Huron in southern Canada. The sequence spans 2.45–2.2 Ga and contains the first unequivocal evidence of glaciation in the form of three layers of diamictites, which are clusters of unconsolidated rock fragments bound in a matrix. Glacial striations and dropstones provide further evidence of glacial origin.

Roscoe observed that rocks containing detrital pyrite and uraninite underlie the lowermost glacial layer. Such deposits are thought to indicate low-oxygen atmospheric conditions. A red bed and oxidized paleosols, which are thought to indicate high-oxygen atmospheric conditions, overlie the uppermost glacial layer, the Gowganda formation (Prasad & Roscoe 1996). Thus, the glaciations and the rise of atmospheric O_2 appear to be related. This timing, of course, makes perfect sense if CH_4 was an important part of the atmospheric greenhouse during

the Late Archean and Early Proterozoic, 2.8–2.3 Ga. The rise of O_2 would have eliminated most of the methane by reducing its photochemical lifetime as well as by constraining the environments in which methanogens could survive.

We would be remiss if we did not point out that two "Snowball Earth" events may have occurred in the Late Proterozoic, around 600 Ma and 750 Ma (Evans, Beukes & Kirshvink 1997, Hoffman et al. 1998, Kirschvink 1992, Kirshvink et al. 2000). In these episodes, Earth's surface may have frozen over entirely to a depth of 1 km or more (Hoffman et al. 1998). Snowball Earth glacial deposits are also overlain in some places by thick layers of carbonate rocks, sometimes called cap carbonates, with unusual textures (indicating rapid deposition) and unusual ratios of carbon isotopes. These cap carbonates may have formed when CO_2 that had built up during the period in which the surface was ice covered was later removed in an episode of rapid carbonate and silicate weathering. In essence, the thought experiment pertaining to the faint young Sun problem described in Section 3.4 appears to have actually occurred on several separate occasions. Although the question of whether the oceans froze over entirely remains controversial (see, e.g., Hyde et al. 2000), the role of the carbonate-silicate cycle during the period of recovery from glaciations is well established. This, again, is something that could also occur on Earth-like extrasolar planets.

4. CHANGES IN ATMOSPHERIC REDOX STATE

Along with the climate, the redox state of a planet's atmosphere influences the possibilities for life on its surface. In particular, multicellular eukaryotic organisms—that is, organisms, such as plants and animals, whose cells have nuclei—require the presence of substantial amounts of molecular oxygen. Hence, much of our interest in how Earth's atmosphere evolved is focused on the rise of atmospheric O_2. This gas is also a biomarker molecule that may eventually be used as evidence of life on extrasolar planets. To make that connection, however, we need to be able to estimate how much O_2 could be generated by strictly abiotic processes. The next two sections describe how that is done.

4.1. Escape of Hydrogen to Space

That terrestrial planets are small enough to lose hydrogen and helium to space is one of the fundamental characteristics distinguishing them from a giant planet. Indeed, their relatively small size is what prevented them from accumulating large amounts of hydrogen and helium in the first place. Thus, we can predict that hydrogen loss is a general characteristic of Earth-sized planets.

Because Earth's hydrogen was not captured gravitationally from the solar nebula, it must have come in as a constituent of condensed materials, which include water ice hydrated silicates, and organic compounds. When hydrogen is lost from such compounds, the remaining part is oxidized. If the hydrogen originated as H_2O, oxygen is left behind. This oxygen can react with reduced materials in Earth's crust

and mantle, or it can accumulate in the atmosphere as O_2. In either case, we need to understand hydrogen escape to estimate how fast oxygen is being produced.

The escape rate of hydrogen from Earth can, in principle, be limited at either of two different heights: the exobase ($\sim$550 km altitude) or the homopause ($\sim$100 km altitude). The exobase is the height at which the atmosphere becomes collisionless. Above this height, the mean free path of the molecules exceeds the local scale height (RT/g). The homopause is the height at which vertical transport becomes dominated by molecular diffusion as opposed to turbulence and mass transport (collectively termed eddy diffusion). Below the homopause, the atmosphere is well mixed; above it, light gases such as hydrogen and helium separate from heavier gases such as N_2 and O_2.

Whether the escape-rate limit is high or low depends on how efficiently hydrogen is lost at the exobase. For present Earth, escape from this level is always efficient. At solar maximum, the exospheric temperature is high (2000–2500 K), and thermal (or Jeans') escape is rapid. The hydrogen atoms that escape are those located in the high-energy tail of the Maxwellian velocity distribution whose velocities exceed the escape velocity ($\sim$11 km/s). At solar minimum, the exospheric temperature is much lower ($\sim$1000 K), so Jeans' escape is inhibited. However, various nonthermal hydrogen-loss processes take up the slack. The most important of these are (*a*) charge exchange between neutral atomic H and the hot H^+ ions trapped in Earth's magnetosphere and (*b*) the polar wind, whereby H^+ ions are accelerated from Earth's polar regions into space along open magnetic field lines (Hunten 1982, Hunten et al. 1989).

As a result of the efficient loss of hydrogen from Earth's upper atmosphere, hydrogen escape is limited by diffusion through the homopause. Physically, this limit exists because the escaping hydrogen experiences friction from the static background atmosphere. The process of diffusion-limited escape is well understood (Hunten 1973, Walker 1977). The escape rate is proportional to the total hydrogen mixing ratio at the homopause (f_{total}):

$$\phi_{\text{escape}} = \left(\frac{b}{H}\right) f_{\text{total}}$$

$$\cong 2.5 \times 10^{13} f_{\text{total}} \quad \text{(H atoms cm}^{-2}\text{ s}^{-1}\text{)}, \tag{3}$$

where b is an average binary diffusion coefficient for the diffusion of hydrogen and H_2 in nitrogen, H is the atmospheric scale height (RT/g), and f_{total} is the sum of the mixing ratios of all the hydrogen-containing species, weighted by the number of hydrogen atoms they contain. Thus,

$$f_{\text{total}} = f(H) + 2f(H_2) + 2f(H_2O) + 4f(CH_4) + \ldots. \tag{4}$$

At the homopause, the dominant hydrogen-bearing species are H and H_2. However, f_{total} remains approximately constant throughout the region of the atmosphere where water vapor does not condense (see Walker 1977). Thus, f_{total} can be conveniently evaluated in the lower stratosphere where measurements can be made

more easily. There, the dominant hydrogen-bearing species are H_2O ($\sim3 \times 10^{-6}$ mixing ratio) and CH_4 (1.6×10^{-6} mixing ratio). Hence, the escape rate is $\sim3.1 \times 10^8$ hydrogen atoms cm^{-2} s^{-1}. This is not very fast by comparison with other geochemical cycles that affect atmospheric oxygen (see below). However, the escape rate would obviously be much higher if reduced gases like CH_4 (or H_2) were more abundant, as they should have been on early Earth, or if the stratosphere was much wetter, as may have been the case on early Venus (Kasting 1988).

4.2. Prebiotic O_2 and H_2 Concentrations

As mentioned earlier, escape of hydrogen to space leads to oxygen accumulation on Earth. Let us compare the hydrogen escape rate above with the rate at which O_2 is generated biologically. If we convert to geochemists' units and divide by 4 for stoichiometry, the hydrogen escape rate just calculated is equivalent to an O_2 production rate of 2.1×10^{10} mol/year. By comparison, the production rate of O_2 from photosynthesis followed by burial of organic carbon and pyrite in marine sediments is $\sim1.8 \times 10^{12}$ mol/year (Holland 2002). Thus, the O_2 production rate from hydrogen escape is slightly greater than 1% of the rate of biological O_2 production. We should be careful here about terminology. Roughly half of the hydrogen that is escaping to space today comes from CH_4, almost all of which is produced biologically. Thus, the actual abiotic O_2 production rate from H_2O photolysis is only approximately half what we have calculated, or 1.0×10^{10} mol/year.

This same methodology can be used to estimate H_2 and O_2 concentrations on prebiotic Earth, as noted originally by Walker (1977). If hydrogen escaping to space were all that was happening, O_2 would accumulate indefinitely in the atmosphere. However, oxygen should also have been consumed by oxidation of Earth's surface and by oxidation of reduced volcanic gases. The most rapid of these two processes, and the one that should have limited atmospheric O_2, is oxidation of volcanic gases. (Surface weathering reactions consume 3–4 times as much O_2 as do volcanic gases today; however, such reactions occur only at relatively high atmospheric O_2 concentrations, whereas volcanic gases can be oxidized at very low O_2 levels.) We can begin by using estimates for the modern rate of volcanic release of gases. According to Holland (2002), the most abundant reduced gas in surface volcanic emissions is H_2, which is being outgassed at a rate of $\sim5 \times 10^{12}$ mol/year, or 3.7×10^{10} hydrogen atoms cm^{-2} s^{-1} (Holland 2002). This corresponds to a potential O_2 sink of half that amount, or 2.5×10^{12} mol/year. This number is 250 times greater than the abiotic O_2 production rate. Hence, we conclude [as did Walker (1977)] that O_2 should not have accumulated in the primitive atmosphere. Instead, H_2 would have accumulated until the escape rate of hydrogen to space was high enough to balance the H_2 flux from volcanos. According to Equations 3 and 4, this balance would have been achieved at an atmospheric H_2 mixing ratio of 7.5×10^{-4}. If we now adjust the volcanic outgassing rate upward by a factor of ~5 to account for higher heat flow on the young Earth, the estimated H_2 mixing ratio increases to a few times 10^{-3}.

In reality, the redox balance of the prebiotic atmosphere would have been more complicated than described above. The hydrogen source would have been augmented by outgassing of other reduced gases such as CO. This interconverts with hydrogen by various photochemically mediated reactions that sum to: $CO + H_2O \leftrightarrow CO_2 + H_2$. Hydrogen should also have been produced (or lost) by rainout of oxidized (reduced) trace gases. For example, rainout of hydrogen peroxide, H_2O_2, would have left H_2 behind, as can be seen by writing the following balanced reaction: $2 H_2O \leftrightarrow H_2O_2 + H_2$. If the H_2O_2 reacted with reduced minerals in the ocean or crust, then this reaction sequence would have constituted a net hydrogen source. Similar redox-balance reactions can be written for atmospheric gases containing sulfur and nitrogen. These rainout terms typically alter predicted atmospheric H_2 mixing ratios by between 10% and 20% (Kasting & Brown 1998).

Once the atmospheric H_2 mixing ratio is known, O_2 can be calculated from a one-dimensional (horizontally averaged) photochemical model, just as O_3 (ozone) is calculated today, given a fixed amount of O_2. A typical result is shown in Figure 7. In this calculation, the surface pressure was fixed at one bar, and the N_2 and CO_2 mixing ratios were fixed at 0.8 and 0.2, respectively. This CO_2 value is approximately the amount needed to compensate for 30%-reduced solar luminosity. The figure shows that some O_2 ($\sim 10^{-3}$ mixing ratio) is indeed present at altitudes
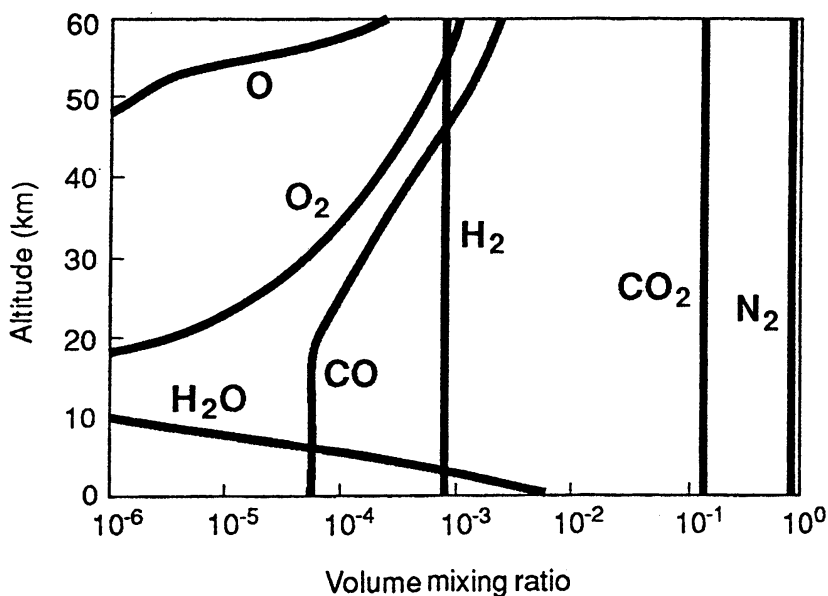

Figure 7 Mixing ratios of major atmospheric species versus altitude for a typical weakly reduced prebiotic atmosphere. The hydrogen abundance is determined by balancing volcanic outgassing with hydrogen escape. O_2 concentrations are predicted using a one-dimensional photochemical model. (From Kasting 1993.)

of 50–60 km. This oxygen is produced mostly from photolysis of CO_2 followed by recombination of O atoms to form O_2. At lower altitudes, however, the O_2 disappears because it is consumed by reactions with volcanic gases and with the CO produced from CO_2 photolysis. These reactions do not occur directly; rather, they are mediated by the photolysis byproducts of H_2O. This chemistry has been explored in detailed photochemical models (Kasting 1979, 1990, 1993; Levine 1982; Pinto, Gladstone & Yung 1980). The resulting ground-level O_2 concentrations are on the order of 10^{-13} PAL (present atmospheric level) (Kasting 1990, 1993). Most of the O_2 near the surface in these latter models is produced by dissociation of CO_2 in lightning discharges: $2CO_2 \leftrightarrow 2CO + O_2$. This again is a transient source of O_2 because the CO produced in this reaction will eventually recombine with oxygen to reform CO_2.

The above discussion of the atmospheric hydrogen budget may appear rather tedious, but it is essential to understanding whether oxygen and ozone are useful biomarkers for the forthcoming TPF and Darwin missions. Models that are not based on a balanced hydrogen budget (e.g., Selsis, Despois & Parisot 2002) can in some circumstances produce high abiotic concentrations of O_2 and O_3 (see their Figures 4 and 5). These authors consciously neglected volcanic outgassing of H_2 in order to calculate an upper limit on oxygen. However, even at low H_2 outgassing rates rainout of photochemically produced oxidants will produce H_2 mixing ratios of the order of 10^{-4} (Kasting et al. 1984) and keep O_2 and O_3 concentrations down to very low values.

4.3. Atmospheric Synthesis of Prebiotic Compounds

Could prebiotic chemical evolution have taken place in such a weakly reduced atmosphere? This question cannot be answered definitively because we do not know for sure how the origin of life occurred. Most biologists would agree, however, that the formation of RNA was a key step in this process. RNA clearly preceded DNA and is capable of catalytic activity (like proteins); hence, it is of interest to determine whether RNA could have been synthesized prebiotically.

A single RNA nucleotide consists of three parts: (*a*) a purine or pyrimidine base containing multiple C–C and C–N bonds, (*b*) a cyclic sugar (ribose) containing C–H and C–O bonds, and (*c*) a phosphate linkage, PO_4^{-3}. The phosphate in RNA (or DNA) derives ultimately from the weathering of rocks. One can safely assume that phosphate was available on primitive Earth, although making it available in a biologically useful form is not necessarily a straightforward task. The sugar ribose ($C_6H_{12}O_6$) can be synthesized by polymerization of formaldehyde (H_2CO). Formaldehyde, in turn, can be formed in the atmosphere from the byproducts of CO_2 and H_2O photolysis (Pinto, Gladstone & Yung 1980). Both CO_2 and H_2O would have been available in the atmosphere of early Earth, so making formaldehyde abiotically should have been relatively easy. Polymerizing it in such a way as to produce predominately ribose, as opposed to other sugars, is a difficult task. We shall assume, however, that nature figured out how to accomplish this.

The most energy-intensive aspect of prebiotic synthesis within the early atmosphere is the production of the nitrogenous base compounds in nucleic acids. The simplest of these is adenine ($C_5H_5N_5$). Adenine is constructed of five molecules of hydrogen cyanide (HCN). Thus, from an atmospheric chemist's standpoint, the key problem is the synthesis of $C \equiv N$ triple bonds.

$C \equiv N$ bonds easily form in an atmosphere that has large amounts of CH_4 and NH_3. In a now-famous experiment, Miller and Urey created spark discharges (simulating lightning) in flasks containing large amounts of these gases and generated a variety of biologically important molecules, including amino acids (Miller 1953, Miller & Urey 1959). However, we now believe that NH_3 was not very abundant because of its high photolysis rate (see Section 3.6). Instead, most of early Earth's nitrogen should have existed as N_2, as it does today (Section 2.2). N_2 is triple-bonded; thus, breaking it apart requires high energies. One way of splitting it is by electric discharge, as in the Miller-Urey experiment. Spark discharge in a mixture containing CH_4 and N_2 produces cyanoacetylene (C_2N) and other $C \equiv N$-bearing compounds (Ferris, Sanchez & Orgel 1968). However, spark discharge in a CO_2-N_2 mixture produces NO rather than HCN (Chameides & Walker 1981, Miller & Schlesinger 1984). The nitrogen atoms produced from splitting N_2 react primarily with oxygen when the atmospheric C:O ratio is less than unity (Chameides & Walker 1981). So, for atmospheres containing more CO_2 than CH_4—which includes most current models of prebiotic Earth—one needs to look for other mechanisms to make HCN.

Zahnle (1986) described one such mechanism. He pointed out that N_2 can be split in Earth's upper atmosphere by a two-step process involving photoionization followed by dissociative recombination:

$$N_2 + h\nu \rightarrow N_2^+ + e$$
$$N_2^+ + e \rightarrow N + N. \tag{5}$$

Some of the nitrogen atoms produced in this manner flow downward, toward the stratosphere where they can react with other molecules or radicals. In atmospheres containing even a few parts per million of CH_4, HCN can be synthesized by the following reactions and by similar reactions involving the methylene radical CH_2:

$$CH_4 + h\nu \rightarrow CH_3 + H$$
$$CH_3 + N \rightarrow HCN + H_2. \tag{6}$$
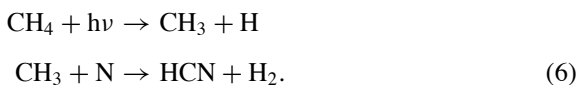
Whether this mechanism could have provided a significant source of HCN depends on how much CH_4 was present in the prebiotic atmosphere. As discussed above, CH_4 is expected to have become quite abundant (hundreds to thousands parts per million) once life, and methanogens, had originated (Section 3.6). What would the CH_4 concentration have been before life arose? The next section explores this question.

4.4. Methane Concentrations in the Prebiotic Atmosphere

Methane is not observed in gases released from surface volcanos (Holland 1978, 1984). The magmas from which surface volcanic gases are released are too oxidized and the pressure is too low for CH_4 to be favored (Holland 1984). The equilibrium is shifted to the left in the reaction $CO_2 + 2H_2O \leftrightarrow CH_4 + 2O_2$, so CO_2 is the dominant carbon-bearing gas released at the surface.

Methane is observed in fluids emanating from hydrothermal vents on the seafloor (Kelley et al. 2001, Welhan 1988). Methane concentrations in high-temperature (300–350°C), axial flow systems like the East Pacific Rise range up to 0.1 μmol/kg (Welhan 1988). This CH_4 is thought to be volcanic (Welhan 1988, Kasting & Brown 1998). Much higher dissolved CH_4 concentrations, 0.2 mmol/kg, have recently been found in cooler (50–60°C) vent fluids emanating from the off-axis Lost City vent field on the Mid-Atlantic ridge (Kelley et al. 2001). Serpentinization of ultramafic rock (peridotite) deep within the oceanic crust is thought to generate this methane. Ultramafic rocks contain high, mantle-like concentrations of (ferrous) iron and magnesium. (The basalt that makes up the seafloor is mafic, but not ultramafic.) Water exposure transforms ultramafic rocks into serpentine minerals, which tend to exclude iron. The iron must then find a stable phase, in this case, magnetite, Fe_3O_4. This mineral contains one Fe^{+2} ion and two Fe^{+3} ions. Because iron is oxidized, something else must be reduced. When pure water is the oxidant, hydrogen is produced: $3FeO + H_2O \rightarrow Fe_3O_4 + H_2$ (Berndt, Allen & Seyfried 1996). When dissolved CO_2 is present in the water, CH_4 is produced instead (Berndt, Allen & Seyfried 1996). The Mid-Atlantic ridge is thought to be a particularly large source of CH_4 because it is a slow-spreading ridge with deep fracture zones that allow penetration of seawater through the basaltic crust to the underlying peridotite (Kelley et al. 2001).

The measurements of dissolved CH_4 in the Lost City vent fluids can be used to estimate a global abiotic source of methane by multiplying the measured concentrations by the total flow of water emanating from off-axis vents on slow-spreading ridges. That flow, in turn, can be estimated from the crustal heat-flow anomaly, that is, the amount of geothermal heat removed by water circulation through the mid-ocean ridges. Mottle & Wheat (1994) estimate a total heat-flow anomaly of 65×10^{18} cal/year, of which $(45–50) \times 10^{18}$ cal/year occurs off axis. This number includes both fast- and slow-spreading ridges. For a vent-fluid temperature of 50°C, the number corresponds to a water flow of $\sim 1 \times 10^{15}$ kg/year. If half of the mid-ocean ridges are slow spreading, and if all water emanating from them has the same dissolved CH_4 concentration as the Lost City vent fluids, then the global CH_4 source from serpentinization of seafloor is 1×10^{11} mol/year. By comparison, the present biological CH_4 source is ~ 535 Tg CH_4/yr, or 3.3×10^{13} mol/year (see Section 3.6). Today, biological methane production outweighs abiotic production by a factor of more than 300. An abiotic methane flux of 1×10^{11} mol/year would produce a CH_4 concentration of ~ 0.5 ppm in a low-O_2 early atmosphere (Pavlov, Kasting & Brown 2001). This estimate assumes a solar extreme ultraviolet (EUV) flux that is higher than today's by a factor of $10^{(200-\lambda)/200}$, where λ is

the wavelength in nanometers. The increase at Ly α (121.6 nm), where most CH_4 photolysis takes place, is thus a factor of $\sim$2.5. This estimate applies to the Sun at $\sim$4.0 Ga, according to data compiled by Walter & Berry (1991).

One might conclude from the above analysis that methane is an excellent biomarker because its concentration on an inhabited planet would be hundreds of times greater than on a lifeless planet. This is not necessarily the case, however, because the abiotic CH_4 source could have been much larger in the past. If the amount of ultramafic rock exposed to aqueous alteration on the seafloor was higher in the past, the methane source from serpentinization reactions could have been higher as well. Today, approximately 21 $\times$ 10^{15} kg of ferrous iron is oxidized to the ferric state during aqueous oxidation of seafloor (Lecuyer & Ricard 1999). The main oxidant today is the sulfate ion SO_4^{-2} in seawater. If, in the distant past, this same amount of ferric iron were produced by serpentinization reactions, the resulting CH_4 source would be $\sim$1.5 $\times$ 10^{12} mol/year. Faster creation of seafloor on a younger, hotter Earth could boost this figure by a factor of five to ten, so an abiotic CH_4 flux of 10^{13} mol/year or more is within the realm of possibility. Evidently, one would need additional biomarker gases to determine whether an observed CH_4-rich atmosphere was being influenced by biology.

4.5. The Rise of Atmospheric Oxygen

At some point in Earth's history, organisms discovered how to perform oxygenic photosynthesis. We can write this reaction as $CO_2 + H_2O \rightarrow CH_2O + O_2$. Here, CH_2O is geochemists' shorthand for more complex forms of organic matter. The oxygen produced from this reaction is responsible for most of the O_2 in Earth's present atmosphere.

We do not have enough space to provide a detailed discussion of the rise of atmospheric oxygen. This topic has been reviewed recently by Catling & Kasting (2003), Holland (1994), and Kasting (1993). Here, we content ourselves with describing two recent advances in our understanding of this problem.

The first is the discovery by Farquhar, Bao & Thiemans (2000) and Farquhar et al. (2001) of mass-independently fractionated sulfur isotopes in Archean rocks. Sulfur has four stable isotopes: ^{32}S, ^{33}S, ^{34}S, and ^{36}S. Historically, only the two most abundant, ^{32}S ($\sim$94% of total sulfur) and ^{34}S (5% of total sulfur), have been analyzed in rocks. These isotopes provide valuable information concerning the antiquity and importance of bacterial sulfate reduction (Canfield, Habicht & Thamdrup 2000, Shen, Buick & Canfield 2001). [Sulfate-reducing bacteria utilize ^{32}S isotope more readily than ^{34}S. Hence, the pyrite (FeS_2) that is produced in sediments by these bacteria is strongly depleted in the heavier isotope.] Farquhar et al. (2000, 2001) looked at ^{33}S and ^{36}S. Figure 8 shows their plot of $\Delta^{33}S$ versus time. $\Delta^{33}S$ represents the deviation in parts per thousand (percentages) from the normal mass fractionation line. In all sulfur-bearing rocks younger than $\sim$2 billion years, the deviation in ^{33}S from its normal ratio to ^{32}S is approximately half that of ^{34}S with respect to ^{32}S. This is because all aqueous chemical (and biological) processes fractionate sulfur in a mass-dependent manner. Mass-independent

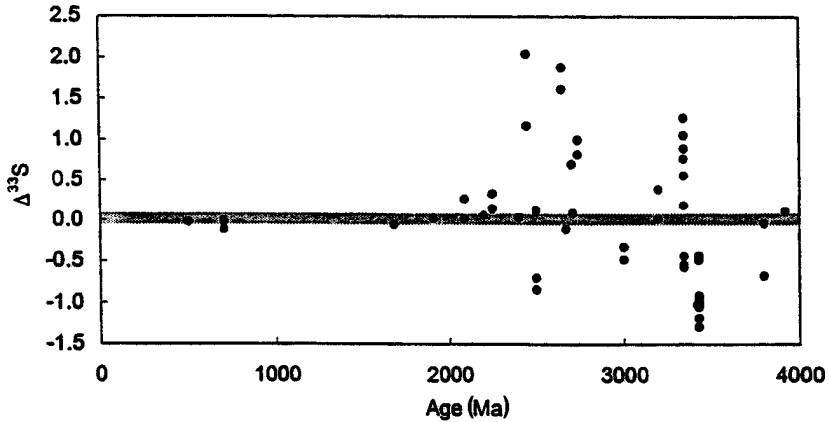**Figure 8** Sulfur-isotope ratios measured in rocks of different ages. Shown on the vertical scale is $\Delta^{33}S$, which represents the deviation of the ^{33}S abundance from its usual relationship to ^{32}S and ^{34}S. Not shown are 73 rock samples of Phanerozoic age (<0.54 Ga), all of which have $\Delta^{33}S$ values equal to $0 \pm 0.1\%$. Photochemical reactions in a low-O_2 atmosphere are required to produce nonzero $\Delta^{33}S$ values. [From Farquhar, Bao & Thiemans (2000). Reprinted with permission of *Science* magazine.]

fractionation (MIF) is thought to originate from gas-phase processes such as photolysis of SO_2 and SO (Farquhar et al. 2001). Both these molecules undergo predissociation at wavelengths between 190 nm and 220 nm. Absorption of photons in this wavelength range results in the creation of excited electronic states that can either deactivate collisionally or decay to separated SO and O (or S and O) atoms or molecules. Subtle quantum mechanical differences in the symmetry of the excited state lead to differences in excited-state lifetimes for molecules containing different sulfur isotopes, and this in turn leads to fractionations that deviate from the normal mass-dependent ratio. (The term mass independent is a misnomer; the fractionations do depend on mass, just not in the normal manner.) Other gas-phase reactions between molecules or radicals may also produce MIF.

Farquhar, Bao & Thiemans (2000) showed that the photolysis of SO_2 in the early atmosphere could be used to set an upper limit on the concentration of O_2. Kasting (2000) and Pavlov & Kasting (2002) extended this argument by pointing out that an even more fundamental requirement for observing MIF is that sulfur must be able to leave the atmosphere in different chemical forms (Figure 9). Today, all sulfur that enters the atmosphere exits as either SO_2 or H_2SO_4 and is eventually combined in the oceanic sulfate reservoir. Hence, any signature of MIF that was produced by atmospheric photochemistry is erased before sulfur enters the sedimentary reservoir. According to computer-modeling experiments performed by Pavlov & Kasting (2002), this remains true at O_2 levels as low as 10^{-5} PAL. Thus, O_2 concentrations prior to ~2 Ga must have been below this level, consistent
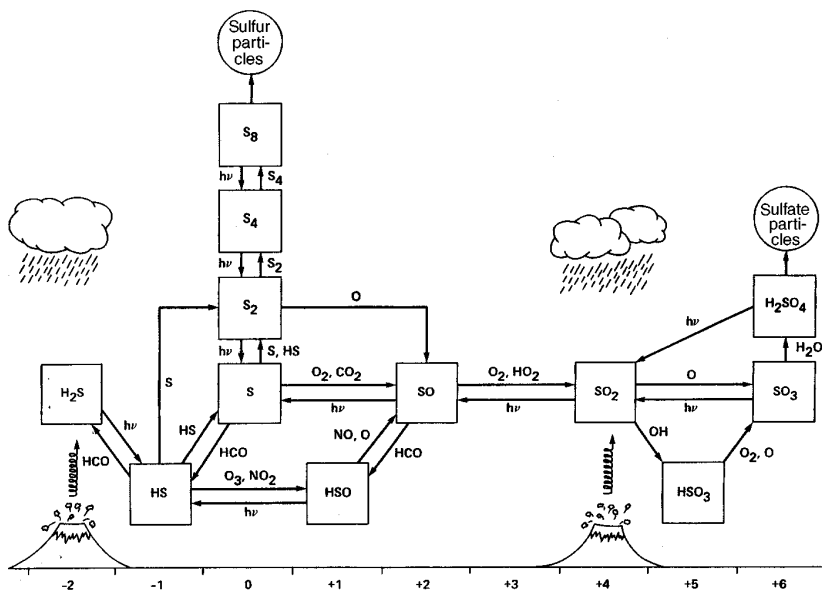
Figure 9 Diagram illustrating sulfur photochemistry in a low-O_2 atmosphere. The horizontal scale is the sulfur-oxidation state. Sulfur is outgassed as either SO_2 or H_2S. It leaves as one of a variety of soluble sulfur gases or as solid S_8 particles. Mass-independent fractionation patterns created by photochemical reactions (e.g., SO_2 photolysis) can be preserved under such circumstances but not in an O_2-rich atmosphere. (From Kasting et al. 1989).

with the model atmosphere shown in Figure 7, in which O_2 virtually disappears below $\sim$20 km altitude. As can be seen in Figure 9, the signature of MIF in sulfur isotopes disappears at approximately the time of the Huronian glaciations, $\sim$2.3 Ga, consistent with the above discussion (Section 3.7) about the glaciations being triggered by the decline of the methane greenhouse.

A second recent development regarding the rise of atmospheric O_2 is the discovery by Brocks et al. (1999) (see also Summons et al. 1999) of organic-biomarker evidence of cyanobacteria and eukaryotes in 2.7 Ga rocks from the Pilbara Craton in Australia. Cyanobacteria are widely believed to have been the first producers of oxygen on our planet. All other photosynthetic algae and higher organisms have ancestors that derived their ability to make oxygen through endosymbiosis with these organisms (Blankenship 1992). Nearly all eukaryotes require O_2 to live. Hence, it appears likely that O_2 was being produced biologically at 2.7 Ga, some 400 million years before atmospheric O_2 levels increased. The apparent time lag between the origin of oxygenic photosynthesis and the rise of O_2 is a topic of current debate (Catling et al. 2001, Holland 2002, Kump, Kasting & Barley 2001). Resolving this issue has relevance for extrasolar-life detection because it would

address whether photosynthesis always results in an immediate rise in atmospheric O_2 concentrations.

4.6. The Rise of Ozone

Accompanying the rise in atmospheric O_2 would have been an increase in stratospheric ozone. Ozone is produced by the reactions

$$O_2 + h\nu \rightarrow O + O$$

$$O + O_2 + M \rightarrow O_3 + M. \tag{7}$$
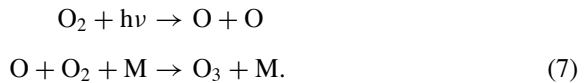
Here, M represents a third molecule that carries away excess energy during the collision of O and O_2. Ozone is important to planetary habitability because it absorbs solar UV radiation in the 200–300 nm region that would otherwise be harmful to organisms (including humans) living at Earth's surface. It is also a potentially observable biomarker in extrasolar planet atmospheres because it has a strong absorption band at 9.6 μm (Angel & Woolf 1996, DesMarais et al. 2002, Leger et al. 1993).

An interesting facet of ozone as a biomarker and as a UV shield is its nonlinear dependence on the abundance of O_2. This was first pointed out by Ratner & Walker (1972). They noted that most O_2 photolysis today occurs above 50 km, where the atmosphere is quite thin. At lower O_2 concentrations, O_2 would be photolyzed below 50 km, where the density is higher and where three-body reactions like the ozone-formation reaction would be faster. Because of this, Ratner & Walker (1972) predicted that the ozone column depth (the integrated amount of ozone in a vertical atmospheric column) would peak at 10^{-3} PAL of O_2. This prediction is not reflected in detailed photochemical models (Kasting, Holland & Pinto 1985, Levine, Hays & Walker 1979) because production of O_3-destroying radicals increases as the ozone layer moves down into the troposphere. Nevertheless, the dependence of O_3 column depth on O_2 abundance remains highly nonlinear (Figure 10).

The consequences of this nonlinear behavior are twofold: (*a*) An effective UV shield was probably established almost as soon as O_2 levels increased, around 2.3 Ga. This would have facilitated the proliferation of eukaryotic phytoplankton. Phytoplankton are photosynthetic organisms that live in the surface ocean, and eukaryotic phytoplankton are much more UV-sensitive than their prokaryotic (nocell nucleus) cousins. A UV shield could have existed earlier as a consequence of absorption by organic-haze particles (Sagan & Chyba 1997), but only if the particle size was smaller or the absorption coefficient was higher than in recent photochemical models (Pavlov, Kasting & Brown 2001). (*b*) Ozone is a better biomarker than O_2 for planets with low but finite (10^{-2}–10^{-1} PAL) O_2 levels. This may well describe Earth during the mid-Protererozoic era, 2.3–0.8 Ga. Thus, for some planets, life detection is easier in the IR spectral region, rather than in the visible.
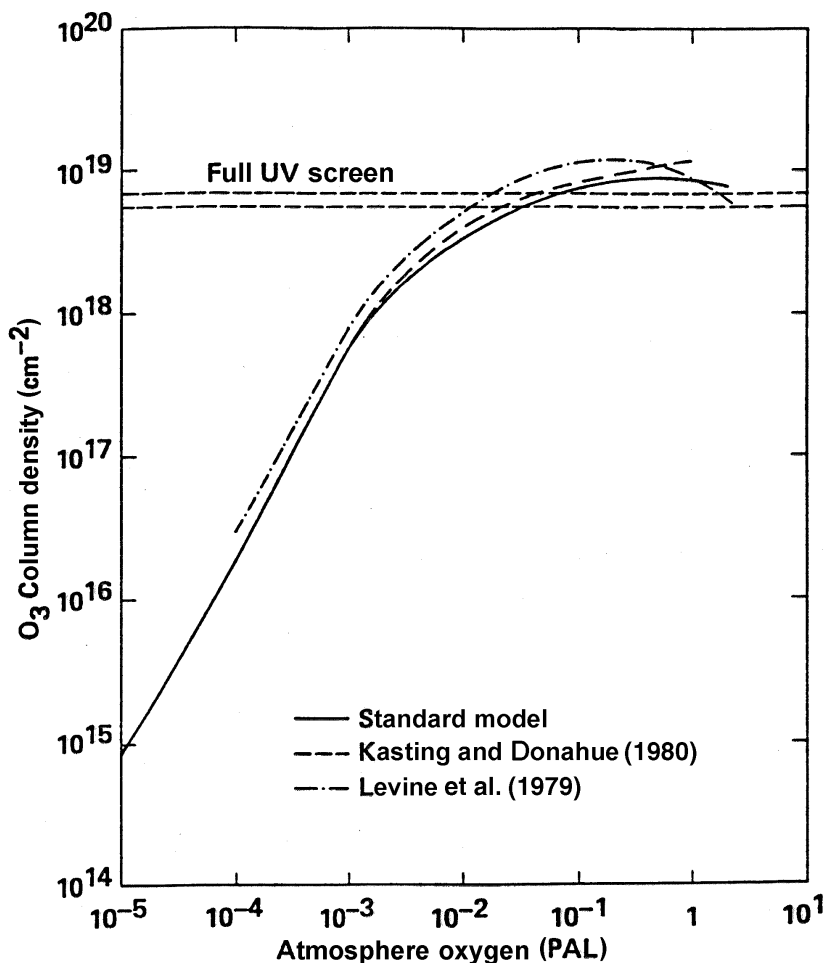
Figure 10 Ozone column depth as a function of atmospheric O_2 mixing ratio. The O_2 scale is in PAL (present atmospheric level). The ozone concentration varies nonlinearly with O_2 as a consequence of the reaction mechanism that produces it. (From Kasting, Holland & Pinto 1985).

5. CONCLUSION

In this brief history of our planet, we focus on factors that make Earth habitable for both simple and complex organisms. Liquid water is a key requirement for both types of organisms, and molecular oxygen is a requirement for complex life, at least as we know it.

As argued in Section 2, most of Earth's water came from farther out in the solar system, most probably from the asteroid belt. This implies that Mars, Earth, and Venus all received substantial endowments of water. However, the delivery of volatiles to inner planets by dynamical scattering in other solar-like planetary systems will likely depend on the orbital parameters and masses of the giant planets in these systems. Keeping water present in liquid form on a planet's surface requires that the planet be situated within the HZ around its star. Current models suggest that HZs are relatively wide for solar-type stars and large planets like Earth that have substantial internal heat sources and, hence, are able to recycle their CO_2 via plate tectonics or other forms of volcanism. Methane was probably a major greenhouse gas on early Earth, after the origin of life and prior to the rise of oxygen. This would likely be true for other inhabited planets as well.

Complex life was enabled by the invention of oxygenic photosynthesis by cyanobacteria, apparently at 2.7 Ga or earlier. Atmospheric O_2 first rose to significant levels around 2.3 Ga, as documented by the disappearance of mass-independent sulfur-isotope fractionation and by other geologic O_2 indicators. The initial rise in O_2 correlates with the Huronian glaciations, suggesting that the O_2 rise eliminated the methane greenhouse that was keeping Earth warm before then. An effective ozone screen developed at the same time, shielding the surface from harmful solar UV radiation and providing a convenient IR biomarker signal that demonstrated to any possible observer that Earth was inhabited. With a little luck and improvements in technology, we should be able to use this same strategy to look spectroscopically for life on distant planets.

The *Annual Review of Astronomy and Astrophysics* is online at
http://astro.annualreviews.org

LITERATURE CITED

Angel JRP, Woolf NJ. 1996. *Sci. Am.* 274:60–66

Bada JL, Bigham C, Miller SL. 1994. *Proc. Natl. Acad. Sci. USA* 91:1248–50

Baross J, Sullivan W, eds. 2003. *Planets and Life: The Emerging Science of Astrobiology.* Cambridge, MA: Cambridge Univ. Press. In press

Beichman CA, Woolf NJ, Lindensmith CA. 1999. *The Terrestrial Planet Finder (TPF): A NASA Origins Program to Search for Habitable Planets (JPL Publ. 99-3).* Pasadena, CA: NASA Jet Propuls. Lab.

Berndt ME, Allen DE, Seyfried WE. 1996. *Geology* 24:351–54

Berner RA, Lasaga AC, Garrels RM. 1983. *Am. J. Sci.* 283:641–83

Borucki WJ, Koch DG, Lissauer JJ, Basri GB, Caldwell JF, et al. 2003. In *Future EUV/UV and Visible Space Astrophysics Missions and Instrumentation*, ed. JC Blades, OHW Siegmund, 4854:129–40. *Proc. SPIE*

Brasier MD, Green OR, Jephcoat AP, Kleppe AK, Van Kranendonk MJ, et al. 2002. *Nature* 416:76–81

Brocks JJ, Logan GA, Buick R, Summons RE. 1999. *Science* 285:1033–36

Bullock MA, Grinspoon DH. 1996. *J. Geophys. Res.* 101:7521–29

Butler P. 2003. See Baross & Sullivan 2003. In press

Caldeira K, Kasting JF. 1992. *Nature* 359:226–28

Cameron AGW. 2000. See Canup & Righter 2000, pp. 133–44

Canfield DE, Habicht KS, Thamdrup B. 2000. *Science* 288:658–61

Canil D. 1997. *Nature* 389:842–45

Canil D. 2002. *Earth Planet. Sci. Lett.* 195:75–90

Canup RM, Agnor CB. 2000. See Canup & Righter 2000, pp. 113–29

Canup RM, Righter K, eds. 2000. *Origin of the Earth and Moon*. Tucson: Univ. Ariz. Press

Catling DC, Kasting JF. 2002. See Baross & Sullivan 2003. In press

Catling DC, Zahnle KJ, McKay CP. 2001. *Science* 293:839–43

Chambers JE, Wetherill GW. 1998. *Icarus* 136:304–27

Chambers JE, Wetherill GW. 2001. *Meteor. Planet. Sci.* 36:381–89

Chameides WL, Walker JCG. 1981. *Orig. Life* 11:291–302

Charbonneau D, Brown TM, Latham DW, Mayor M. 2000. *Ap. J.* 529:L45-49

Chyba CF. 1989. *Nature* 343:129–32

Cooney CL. 1975. *Biotechnol. Bioeng.* 17:1119–35

Corliss JB, Baross JA, Hoffman SE. 1981. *Ocean Acta* 4(Suppl.):59–69

Craig H, Clarke WB, Beg MA. 1975. *Earth Planet. Sci. Lett.* 26:125–32

Delano JW. 2001. *Orig. Life Evol. Biosph.* 31:311–41

Delsemme AH. 2001. *Am. Sci.* 89:432–42

Des Marais DJ, Harwit MO, Jucks KW, Kasting JF, Lin DNC, et al. 2002. *Astrobiology* 2:153–81

Dole SH. 1964. *Habitable Planets for Man*. New York: Blaisdell

Evans DA, Beukes NJ, Kirshvink JL. 1997. *Nature* 386:262–66

Farquhar J, Bao H, Thiemans M. 2000. *Science* 289:756–58

Farquhar J, Savarino J, Airieau S, Thiemens MH. 2001. *J. Geophys. Res.* 106:1–11

Fedo CM, Whitehouse MJ. 2002. *Science* 296:1448–52

Ferris JP, Sanchez RA, Orgel LE. 1968. *J. Mol. Biol.* 33:693–704

Forget F, Pierrehumbert RT. 1997. *Science* 278:1273–76

Gough DO. 1981. *Sol. Phys.* 74:21–34

Halliday AN, Lee D-C, Jacobsen SB. 2000. See Canup & Righter 2000, pp. 45–62

Halliday AN, Lee D-C, Porcelli D, Schonbachler M, Rehkamper M. 2001. *Philos. Trans. R. Soc. London Ser. A* 359:2111–34

Hart MH. 1978. *Icarus* 33:23–39

Hart MH. 1979. *Icarus* 37:351–57

Hartmann WK, Phillips RJ, Taylor GJ. 1986. *Origin of the Moon*. Houston: Lunar Planet. Inst.

Hoffman PF, Kaufman AJ, Halverson GP, Schrag DP. 1998. *Science* 281:1342–46

Holland HD. 1962. In *Petrologic Studies: A Volume to Honor A.F. Buddington*, ed. AEJ Engel, HL James, BF Leonard, pp. 447–77. New York: Geol. Soc. Am.

Holland HD. 1978. *The Chemistry of the Atmosphere and Oceans*. New York: Wiley

Holland HD. 1984. *The Chemical Evolution of the Atmosphere and Oceans*. Princeton: Princeton Univ. Press

Holland HD. 1994. In *Early Life on Earth*, ed. S Bengtson, pp. 237–44. New York: Columbia Univ. Press

Holland HD. 2002. *Geochim. Cosmochim. Acta* 66:3811–26

Houghton JT, Meira Filho LG, Bruce JP, Lee H, Callander BA, Haites EF, eds. 1994. *Climate Change, 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emission Scenarios*. Cambridge, MA: Cambridge Univ. Press

Hoyle F. 1957. *The Black Cloud*. New York: Harper

Holyle F. 1958. In *Stellar Populations*, ed. DJK O'Connell. Vatican City: Specola Vaticana

Hunten DM. 1973. *J. Atmos. Sci.* 30:1481–94

Hunten DM. 1982. *Planet. Space Sci.* 30:773–83

Hunten DM, Donahue TM, Walker JCG, Kasting JF. 1989. In *Origin and Evolution of Planetary and Satellite Atmospheres*, ed. SK Atreya, JB Pollack, MS Matthews. Tucson: Univ. Ariz.

Hyde WT, Crowley TJ, Baum SK, Peltier WR. 2000. *Nature* 405:425–29

Joshi MM, Haberle RM, Reynolds RT. 1997. *Icarus* 129:450–65

Kasting JF. 1990. *Orig. Life* 20:199–231

Kasting JF. 1991. *Icarus* 94:1–13

Kasting JF. 1993. *Science* 259:920–26

Kasting JF. 1982. *J. Geophys. Res.* 87:3091–98

Kasting JF, Brown LL. 1998. In *The Molecular Origins of Life: Assembling the Pieces of the Puzzle*, ed. A Brack, pp. 35–56. New York: Cambridge Univ. Press

Kasting JF, Eggler DH, Raeburn SP. 1993. *J. Geol.* 101:245–57

Kasting JF, Holland HD, Pinto JP. 1985. *J. Geophys. Res.* 90:10497–510

Kasting JF, Pavlov AA, Siefert JL. 2001. *Orig. Life Evol. Biosph.* 31:271–85

Kasting JF, Pollack JB, Crisp D. 1984. *J. Atmos. Chem.* 1:403–28

Kasting JF, Toon OB, Pollack JB. 1988. *Sci. Am.* 256:90–97

Kasting JF, Whitmire DP, Reynolds RT. 1993. *Icarus* 101:108–28

Kasting JF, Zahnle KJ, Pinto JP, Young AT. 1989. *Orig. Life* 19:95–108

Kelley DS. 2001. *Nature* 412:145–49

Kirschvink JL. 1992. In *The Proterozoic Biosphere: A Multidisciplinary Study*, ed. JW Schopf, C Klein, pp. 51–52. Cambridge, MA: Cambridge Univ. Press

Kirshvink JL, Gaidos EJ, Bertaini LE, Benkes NJ, Gutzmer J, et al. 2000. *Proc. Natl. Acad. Sci. USA* 97:1400–5

Knoll AH, Barghoorn ES. 1977. *Science* 198:396–98

Kral TA, Brink KM, Miller SL, McKay CP. 1998. *Orig. Life Evol. Biosph.* 28:311–19

Kuhn WR, Atreya SK. 1979. *Icarus* 37:207–13

Kump LR, Kasting JF, Barley ME. 2001. *Geol. Geochem. Geophys. No. 2000 GC 000114*

Lange MA, Ahrens TJ. 1982. *Icarus* 51:96–120

Laskar J, Robutel P. 1993. *Nature* 361:608–14

Lecuyer C, Ricard Y. 1999. *Earth Planet. Sci. Lett.* 165:197–211

Leger A. 2000. *Adv. Space Res.* 25:2209–23

Leger A, Pirre M, Marceau FJ. 1993. *Astron. Astrophys.* 277:309–13

Levine JS. 1982. *J. Mol. Evol.* 18:161–72

Levine JS, Hays PB, Walker JCG. 1979. *Icarus* 39:295–309

Maher KA, Stevenson DJ. 1988. *Nature* 331:612–14

Marcy GW, Butler RP. 1998. *Annu. Rev. Astron. Astrophys.* 36:57–97

Matsui T, Abe Y. 1986. *Nature* 319:303–5

Matsui T, Abe Y. 1986. *Nature* 322:526–28

McKay CP, Pollack JB, Courtin R. 1991. *Science* 253:1118–21

Melosh HJ, Vickery AM. 1989. *Nature* 338:487–89

Miller SL. 1953. *Science* 117:528–29

Miller SL, Schlesinger G. 1984. *Orig. Life* 14:83–90

Miller SL, Urey HC. 1959. *Science* 130:245–51

Mischna MM, Kasting JF, Pavlov AA, Freedman R. 2000. *Icarus* 145:546–54

Mojzsis SJ, Arrhenius G, McKeegan KD, Harrison TM, Nutman AP, Friend CRL. 1996. *Nature* 384:55–59

Mojzsis SJ, Harrison TM, Pidgeon RT. 2001. *Nature* 409:178–81

Morbidelli A, Chambers J, Lunine JI, Petit JM, Robert F, et al. 2000. *Meteor. Planet. Sci.* 35:1309–20

Morbidelli A, Petit J-M, Gladman B, Chambers J. 2001. *Meteor. Planet. Sci.* 36:371–80

Mottl MJ, Wheat CG. 1994. *Geochim. Cosmochim. Acta* 58:2225–37

Newsom HE, Jones JH, eds. 1990. *Origin of the Earth*. New York: Oxford Univ. Press

Owen T, Bar-Nun A, Kleinfeld I. 1992. *Nature* 358:43–46

Pace NR, Olsen GJ, Woese CR. 1986. *Cell* 45:325–26

Pavlov AA, Kasting JF. 2002. *Astrobiology* 2:27–41

Pavlov AA, Kasting JF, Brown LL. 2001. *J. Geophys. Res.* 106:23267–87

Pavlov AA, Kasting JF, Brown LL, Rages KA, Freedman R. 2000. *J. Geophys. Res.* 105:11981–90

Pavlov AA, Kasting JF, Eigenbrode JL, Freeman KH. 2001. *Geology* 29:1003–6

Pepin RO. 1991. *Icarus* 92:2–79

Petit JM, Morbidelli A, Valsecchi GB. 1999. *Icarus* 141:367–87

Pinto JP, Gladstone CR, Yung YL. 1980. *Science* 210:183–85

Prasad N, Roscoe SM. 1996. *Catena* 27:105–21

Prinn RG, Fegley B. 1989. In *Origin and Evolution of Planetary and Satellite Atmospheres*, ed. SK Atreya, JB Pollack, MS Matthews, pp. 78–136. Tucson: Univ. Ariz. Press

Ratner MI, Walker JCG. 1972. *J. Atmos. Sci.* 29:803–8

Righter K, Drake MJ. 1997. *Earth Planet. Sci. Lett.* 146:541–53

Righter K, Drake MJ. 1999. *Earth Planet. Sci. Lett.* 171:383–89

Robert F. 2001. *Science* 293:1056–58

Roscoe SM. 1969. *Geol. Surv. Can. Spec. Pap. 68-40*. Toronto: Can. Geol. Surv. 205 pp.

Roscoe SM. 1973. *Geol. Soc. Can. Spec. Pap.* 12:31–48. Toronto: Can. Geol. Surv.

Rosing MT. 1999. *Science* 283:674–76

Rossow WB, Henderson-Sellers A, Weinrich SK. 1982. *Science* 217:1245–47

Rubey WW. 1955. In *Crust of the Earth*, ed. A Poldervaart, pp. 631–50. New York: Geol. Soc. Am.

Rye R, Kuo PH, Holland HD. 1995. *Nature* 378:603–5

Safronov VS. 1969. *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets*. Transl. M Nanka, 1972 as NASA TT F-677

Sagan C, Chyba C. 1997. *Science* 276:1217–21

Sagan C, Mullen G. 1972. *Science* 177:52–56

Schopf JW. 1993. *Science* 260:640–46

Schwarzschild M. 1958. *Structure and Evolution of the Stars*. Princeton: Princeton Univ. Press

Selsis F, Despois D, Parisot J-P. 2002. *Astron. Astrophys.* 388:985–1003

Shen Y, Buick R, Canfield DE. 2001. *Nature* 410:77–81

Sleep NH, Zahnle KJ, Kasting JF, Morowitz HJ. 1989. *Nature* 342:139–42

Stevenson DJ. 1983. In *Earth's Earliest Biosphere: Its Origin and Evolution*, ed. JW Schopf, pp. 32–40. Princeton: Princeton Univ. Press.

Summons JR, Jahnke LL, Hope JM, Logan GA. 1999. *Nature* 400:554–57

van Zuilen MA, Lepland A, Arrhenius G. 2002. *Nature* 418:627–30

Walker JCG. 1977. *Evolution of the Atmosphere*. New York: Macmillan

Walker JCG. 1990. *Palaeogeogr. Palaeoclim. Palaeoecol.* 82:261–89

Walker JCG, Hays PB, Kasting JF. 1981. *J. Geophys. Res.* 86:9776–82

Walter FM, Barry DC. 1991. In *The Sun in Time*, ed. CP Sonett, MS Giampapa, MS Matthews, pp. 633–57. Tucson: Univ. Ariz. Press

Ward PD, Brownlee D. 2000. *Rare Earth: Why Complex Life is Uncommon in the Universe*. New York: Copernicus

Welhan JA. 1988. *Chem. Geol.* 71:183–98

Wetherill GW. 1985. *Science* 228:877–79

Wilde SA, Valley JW, Peck WH, Graham CM. 2001. *Nature* 409:175–78

Williams DM, Kasting JF, Caldeira K. 1996. In *Circumstellar Habitable Zones*, ed. LR Doyle, pp. 43–62. Menlo Park, CA: Travis House

Woese CR, Fox GE. 1977. *Proc. Natl. Acad. Sci. USA* 74:5088–90

Yin Q, Jacobsen SB, Yamashita K, Blichert-Toft J, Telouk P, Albarede F. 2002. *Nature* 418:949–52

Zahnle KJ. 1986. *J. Geophys. Res.* 91:2819–34

Zahnle KJ, Kasting JF, Pollack JB. 1990. *Icarus* 84:502–27

Zahnle KJ, Sleep N. 1997. In *Comets and the Origin and Evolution of Life*, ed. PJ Thomas, CF Chyba, CP McKay, pp. 175–208. New York: Springer

Zahnle KJ, Walker JCG. 1982. *Rev. Geophys. Space Phys.* 20:280–92

Annual Review of Astronomy and Astrophysics
Volume 41, 2003

Contents

ERRATA

An online log of corrections to *Annual Review of Astronomy
and Astrophysics* chapters (if any, 1997 to the present)
may be found at http://astro.annualreviews.org/errata.shtml