CONVERSATIONS ON THE HABITABILITY OF WORLDS:
THE IMPORTANCE OF VOLATILES

J.-L. BERTAUX (jean-loup.bertaux@aerovre.jussieu.fr)
Service d’Aéronomie, CNRS, Paris, France

M. CARR (carr@usgs.gov)
USGS, Menlo Park, California USA

D. J. DES MARAIS (david.j.desmarais.nasa.gov)
NASA Ames Research Center, Moffett Field, California USA

E. GAIDOS (gaidos@hawaii.edu)
University of Hawaii, Honolulu, Hawaii USA

Received: February 14, 2006; Accepted in final form: April 10, 2007

Abstract. Our scientific forefathers discuss the interrelationships between water, climate, the atmosphere, and life on Earth and other terrestrial planets at a workshop in Nichtchâtel, Switzerland.

Guillaume: Good morning everyone. My name is Charles Guillaume and I am very honored to chair this session on “Planetary Volatiles, Atmospheres, and Habitability”. I look forward to what our distinguished session speakers have to say about planetary habitability. I must confess that I know nothing of the subject, having spent my entire career working on the properties of nickel-steel alloys that have application in precision instruments. I thought this was a worthwhile endeavor, and evidently so did the Norwegian Nobel Committee in 1920. A reminder that our format will consist first of four talks, followed by a panel discussion involving our speakers. We will first hear from Charles Messier, who will speak on “The Origin of Earth’s Water”, followed by William Herschel, who will discuss “The History of Water and Climate on Mars”. We then have Charles Lyell, who will present “Volatile Reservoirs and Exchange Processes on Earth and Mars”, and finally Charles Darwin, who will finish with “Co-evolution of the Atmosphere and Life on Earth”. Well, we must stay on schedule; this is Switzerland, after all, so I would like to introduce the first speaker, Charles Messier. Mssr. Messier was a self-educated astronomer until he began working for the French navy in 1751. Six years later he began searching for Comet Halley, which was predicted to return in 1758. His independent recovery of that famous object in early 1759 launched his career as a comet hunter, but he is best known for the catalog of nebulae that he developed while searching for comets. He was elected to the Royal Society of London in 1764, and to the Royal Academy of Sciences of Paris in 1770. He was appointed Astronomer of the Navy the following year. Charles?
1. The Origin of Earth’s Water

Messier: Merci. May I have the first slide? All life on Earth requires water; it seems to be an immutable requirement of planetary habitability, at least for life as we know it. Therefore, what is the origin of Earth’s water? Because Earth is not alone in this region of the solar system, we should also consider the case of Venus and Mars, where liquid water is or could have been present and where life might have emerged independently. Also, if we find a good theory for the origin of water on Earth, it must also be compatible with the current content of water on the two other planets.

In planetary environments, water typically exists in the following four forms: solid ice, liquid, vapor, and as water molecules that are bound in a great variety of minerals, for example, gypsum: $\text{CaSO}_4 \cdot 2\text{(H}_2\text{O)}$. These minerals can liberate their water when they are heated, which indeed occurred during the formation of the planets. Later, radioactive heating and outgassing from the interiors of planets transferred this water to the oceans and atmosphere.

1.1. Planetary Inventories of $\text{H}_2\text{O}$

The present inventory of water on the three planets is as follows. On Venus, where the crust is hot and degassed, the only reservoir is in the atmosphere as vapor. If transformed into liquid, it would represent a layer of only 3 cm. This estimate may be calculated from the $\text{H}_2\text{O}$ vertical profile data obtained by Moroz et al. (1983), re-interpreted by Ignatiev et al. (1997) as a constant mixing ratio of 3 ppmv. Let us take 3 cm as representative for Venus, which makes it the driest, by far, of the three planets today. The Earth has 2.8 km of liquid water, if spread uniformly over the surface, and an additional amount of chemically and physically bound water in the crust and mantle of about 4 km equivalent thickness. It is interesting that, if the Earth had more liquid water, a total of about 11 km or more equivalent thickness, then there would be no land. The Himalaya summits would be submerged.

What is the quantity of $\text{H}_2\text{O}$ on Mars now? The main measurable reservoir is the permanent North polar cap; its volume was measured by Mars Orbiter Laser Altimeter (MOLA) (Zuber et al., 1998). This quantity of ice corresponds to a layer of about 9 m, if spread uniformly as liquid on the surface. The Mars Express orbiter discovered that the permanent south polar cap is made mainly of $\text{H}_2\text{O}$ ice, with a layer of $\text{CO}_2$ ice on top. The visible south polar cap is much smaller than its northern counterpart, but overlays ice-rich layered deposits that rival its northern counterpart. Neutron and gamma ray measurements also discovered a large fraction of $\text{H}$ in the surface at high latitudes, but these techniques can probe only the first meter of the subsurface. Therefore, the total amount of measured $\text{H}_2\text{O}$ reservoir today is about 12 m ($\pm$2 m). This corresponds to a volume of $1.7 \times 10^6$ km$^3$ of water, a number to remember when William Herschel will present later the geological
evidence for erosion on the surface, and estimate the amount of water necessary to create this erosion.

1.2. **Evolution of Planetary Volatiles**

Coming back to Earth, do we know the origin of its plentiful water? Well, to know better, we have to journey backward in time to the formation of the solar system, 4.55 billion years ago. And we have to combine two scientific disciplines that fortunately are communicating with each other better and better, although each of them requires very specialized skills. One discipline performs ever more refined measurements of elemental and isotopic compositions of Earth, Moon, and meteorites, including meteorites derived from Mars and dust from comets. This allows us to reconstitute a remarkably detailed and accurate chronology of a number of events in the early history of the solar system. The other scientific discipline develops theories of the formation of planets through accretion of material from the protosolar nebula. Numerical simulations of mutual interactions between planet-forming objects play an important role in model development.

In the early descriptions by Safronov (1969), small dust particles aggregated into km-size bodies, called planetesimals, that eventually assembled into the planetary bodies, such as the terrestrial planets, that assumed more or less their present orbital location and present sizes. The giant planets accreted an additional huge amount of primitive nebula gas. One problem with this picture is that the study of meteorites indicates that material formed inward of 2.5-3 AU (Astronomical Unit, equal to the Sun-Earth distance) was too dry to supply the quantity of water in Earth’s oceans (Lunine et al., 2003).

However, the theory of accretion has progressed since the early picture of Safronov. After the first stage of formation of planetesimals, dynamical simulations indicate that these planetesimals combined to form Mercury- to Mars-mass planetary embryos in quasi-circular and coplanar orbits, from 0.3 to ~4 AU (Petit et al., 2000). These orbits extended outward to the asteroid belt, where the original material contained minerals with bound water. Minerals such as these have survived in a special class of meteorites, the carbonaceous chondrites.

Then, Jupiter accreted its gaseous envelope and reached its present mass, with enormous consequences for the planetary embryos. In particular, the ones orbiting at more than 2 AU and nearest the orbit of Jupiter, were either ejected from the solar system, sent to the Sun, or sent to orbits crossing the orbits of the growing terrestrial planets, provoking gigantic collisions, but providing a good supply of water to the otherwise dry planetary embryos (Lunine et al., 2003). Interestingly, these and other simulations do not usually produce a Mars-mass planet in an orbit similar to the Mars orbit; they are either much more massive in a gas-free system, or, with gas drag, a number of very small planets form.

Venus and Earth were therefore growing through these gigantic collisions with embryos sent by Jupiter from orbits beyond that of Mars. The development of this
scenario was stimulated by the work of Cameron and Benz (1991), who described a theory that the Moon was formed through a gigantic impact of the Earth with a Mars-size body, now called Theia in the literature. Though, by principle, we do not like to invoke catastrophic events to explain what we see (and in particular our own existence), it seems that such a catastrophe is the only way to explain the present composition of Earth and Moon, as inferred from samples returned by the Apollo missions. This theory is now widely accepted, and the collision event is now dated to have occurred rather precisely, at 40 to 50 million years after the formation of the oldest meteorites, 4.567 billion years ago.

But what happened to water during this terrible collision? It is easy to imagine that the oceans definitively could have been lost to space. Detailed simulations of collisions show a more complicated picture, however. If the Earth had only an atmosphere and no oceans (water present only as steam), then 90% of the atmosphere would survive a moderate velocity collision (Genda and Abe, 2004). It is reasonable to assume that the collision occurred at a moderate velocity because, otherwise, the Earth would have decreased in size and the Moon would not have formed after the collision with Theia. But if there was an ocean (if water had already condensed before 40 million years after accretion), then the mechanics of the shock are entirely different. The transmission of the shock through the liquid phase is easier, the resulting velocity of the shocked atmosphere is larger, even at antipode of the shock, and the ocean might have been partially lost to space and the atmosphere totally lost (Genda and Abe, 2005).

If such an event indeed occurred, it could have lowered the content of $^{36}$Ar (a noble, non-radiogenic gas) in Earth’s atmosphere to a quantity 50 times less than that on Venus, as is found today. This difference has been a long standing-puzzle. If water existing only as steam on Venus early in its history, then its primitive atmosphere, including original inventory of $^{36}$Ar, would have mostly survived the large collisions. Pursuing this scenario further, we could very well imagine that, indeed, the gigantic collisions with the Earth could have also eliminated a substantial fraction of the ocean. Some scientists believe that the challenge to understanding the history of water on Earth is not to find an adequate source, because the sources are large, but rather to find ways to eliminate a good fraction of its water. Maybe the collision with Theia, or another large collision that preceded it, could have eliminated 80% of the water. Without this enormous loss of water, the ocean would have inundated the Himalaya. Possibly life could still emerge on this Planet Ocean, but we would not be here to speculate on the habitability of terrestrial planets.

Studies of deuterium to hydrogen ratios (D/H) in the atmospheres (and in the ocean) of terrestrial planets and in meteorites and comets also offer clues about the origin of water and its evolution through the aeons (Robert, 2001). The terrestrial D/H value is 149 parts per million, similar to that of seawater that is referred to as SMOW (Standard Mean Ocean Water). It is very near the values commonly found in carbonaceous chondrites, whereas D/H values found in the atmospheres of three long-period comets are twice as great (Balsiger et al., 1995; Bockelée-
Such differences in D/H do not favor comets as the main source of water on Earth. One may, however, claim that these three comets are not representative, or invoke another, *ad hoc* source of water with less deuterium that would combine to yield the present D/H ratio in the oceans (Owen and Bar-Nun, 2000).

Over time the atmosphere of a terrestrial planet may be enriched in D through differential escape. Water vapor is transported upward and photo-dissociated by solar UV radiation to yield H and D atoms that diffuse up to the exosphere, the most external part of an atmosphere, where the density is so low that essentially no collisions occur. Atoms having velocities larger than the escape velocity are sent on hyperbolic trajectories and lost from the planet. Clearly the higher mass of D atoms makes their escape less likely, with a subsequent enrichment of HDO, relative to H$_2$O, in the atmosphere. The escape of D atoms may be further reduced by preferential condensation of HDO as ice, preventing much of the HDO from attaining higher altitudes where it is photodissociated by solar UV (Bertaux and Montmessin, 2001).

The atmosphere of Venus is enormously enriched in HDO; its HDO/H$_2$O value is \(\sim 130\) times that of SMOW. If no D atoms escaped to space during the entire history of Venus after the end of big collisions that provoked undifferentiated hydrodynamic escape to space, the atmosphere's present-day HDO content would indicate that there was never more than \(3 \times 130\) cm, or 4 m of water. This is a very small quantity indeed, compared to Earth. But this is an absolute lower limit; if there is now some escape of D atoms (and we will know with the forthcoming ESA Venus Express mission), then some scaling to the present escape of H can be done and extrapolated back in time for a more accurate estimate of past water inventories on Venus. For Earth, we assume that the escape of hydrogen to space caused little or no enrichment in the D/H values of the oceans and atmosphere. As a matter of fact, the present measured escape rate of H of \(2 \times 10^8\) atoms cm$^{-2}$ sec$^{-1}$ represents a loss of only 4 m of liquid water during all of Earth’s history (Bertaux, 1974). On Mars, the HDO enrichment is found to be \(\sim 5.5\) SMOW. Applying the same line of reasoning as for Venus, we infer that Mars once had \(\sim 70\) m of water, or \(10^7\) km$^3$. Is this quantity sufficient to explain all of the water erosion features that we observe on Mars? This inventory seems to be sufficient, especially when considering that liquid water may have flowed many times, either through seasonal cycles or through climatic cycles controlled by the rotation axis inclination and orbit eccentricity. In this case, there might be no significant amount of water ice buried below the surface of Mars. But William Herschel will now give the facts about what we have actually observed on Mars.

**Guillame:** Thank you, Charles. Our next speaker is William Herschel. Sir Herschel began his career as a musician - an organist, I believe - first in Germany and then in England. In 1773, At the age of 35 he became interested in astronomy and started constructing the world’s most advanced telescopes. Eight years later he discovered
the planet Uranus and was soon after elected to the Royal Society as well as appointed Court Astronomer. Besides his work on nebulæ, he also discovered two moons of Uranus, two moons of Saturn, and he first proposed that the polar caps of Mars were made of ice and snow. William?

2. The History of Water and Climate on Mars

**Herschel:** Thanks. I will discuss the hydrologic history of Mars and what it means for climate on that planet. Although the origin of the valleys and channels on Mars has been debated for over thirty years their origin remains almost as puzzling as it was when they were first observed in 1972 during the Mariner 9 mission. For liquid water to be thermodynamically stable, temperatures must exceed 273 K and the atmospheric pressure must exceed 6.1 mbar. For most of the planet the total pressure is less than 6.1 mbar and liquid water at the surface would rapidly boil and freeze. Ice, on heating, would sublimate rather than pass through an intervening liquid phase. Liquid water is stable at the pressures and temperatures found at low elevations at midday in summer. However, the partial pressure of water vapor in the atmosphere typically falls 2-3 orders of magnitude short of 6.1 mbar and thus any liquid water would eventually evaporate. Liquid water could exist transitly in the upper centimeter of an ice-rich soil heated by the Sun, if water vapor were inhibited from diffusing into the atmosphere, but the amount of water involved would be minute. These conditions were mostly understood when the valleys and channels were first discovered, so that origins other than water erosion were explored. Included were faulting, mass wasting, and erosion by wind, lava, liquid carbon dioxide, and other exotic fluids. Despite these possibilities, the broad consensus is that most of the channels and valleys were cut by water. The consensus has been reinforced recently by detection of water soluble minerals at the surface in several places (Gendrin et al., 2005), by discovery of evaporites in water-lain sediments at Meridiani (Squyres et al., 2004), by extensive aqueous alteration of rocks in the Columbia Hills of Gusev Crater (Squyres et al., 2006), and by confirmation that water-ice is widespread just below the surface at high latitudes (Boynton et al., 2002; Mitrofanov et al., 2002; Feldman et al., 2004). In the following discussion we will assume that, except for lava channels in volcanic regions and some local mass-wasting and ice sculpted features, most of the valleys, channels, and gullies are water-worn.

2.1. Stability of Liquid Water

With mean annual temperatures close to 215 K at the equator and 160 K at the poles, the ground is everywhere frozen to kilometer depths. The exact depth at which liquid water might be found depends on the heat flow, the thermal conductivity of the crustal rocks, and the salinity of the water. Taking plausible ranges
for these parameters Clifford (1993) and Clifford and Parker (2001) estimated the mean thickness of the cryosphere (permanently frozen crust) to be in the 2-20 km range. In volcanically active areas, the thickness of the cryosphere could be reduced to zero as hydrothermal fluids reach the surface, but no such areas, either fossil or active, have been detected. The thickness range applies not only to today but probably to much of post-Noachian times since cold climatic conditions appear to have prevailed for much of Mars’ history, and the increase in heat flow back in time (McGovern et al., 2002) is unlikely to result in values that fall outside the range considered by Clifford and co-workers. (The Noachian is the earliest epoch in Mars history ending sometime between 3.8 and 3.5 billion years ago.)

Under present conditions, not only is liquid water unstable everywhere on the surface, but so is ice, since the surface temperature everywhere exceeds the frost point at some time during the year. Water-ice is present at the poles because sublimation losses during summer are balanced by winter accumulation. At latitudes higher than 40 degrees, ice, while unstable on the surface, is stable at depths greater than roughly one meter where the mean temperature of the ground never exceeds the frost point. Consistent with these relations, hydrogen in amounts equivalent to several tens of percent ice, has been detected below a thin dehydrated layer at high latitudes (Boynton et al., 2002; Mitrofanov et al., 2002; Feldman et al., 2004). In contrast, at low latitudes temperatures exceed the frost point at all depths so that ground ice is unstable: Any ice present will eventually sublimate and the water vapor lost to the atmosphere. Detection of several percent water at low latitudes (Feldman et al., 2004) is, therefore, inconsistent with the stability relations. The water detected may be chemically bound water, or water inherited from an earlier era when conditions were different, and has yet to completely diffuse from the regolith.

The conditions just described are for the present day. The stability of water near the surface is sensitive to the obliquity, and Mars may have recently emerged from an era when its obliquity was higher (Laskar et al., 2002). At an obliquity lower than today’s, equatorial temperatures rise and the amount of water in the atmosphere probably falls as the water-ice caps grow. As a result the latitude belt over which ice is unstable at all depths widens (Mellon and Jakosky, 1995). At a higher obliquity the reverse occurs. Equatorial temperatures fall and the amount of water present in the atmosphere increases and with it the frost point temperature. Ice may then be stable at shallow depths below the surface at all latitudes. At the highest values of obliquity water ice may be driven from the poles and accumulate on the surface at low to mid latitudes.

Given the stability conditions just described, a major issue with respect to the seemingly water worn features is whether climate changes were required to form them and, if so, when they occurred, what their magnitude was, how sustained the climatic changes were, and what caused them. Coupled to these issues are others such as the former presence of oceans and the fate of the missing water discussed by Mssr. Messier. These issues will be addressed by examining the different types
of water worn features and what they might imply about the hydrologic history of the planet.

Three classes of likely water worn features are recognized. Outflow channels are linear swaths of scoured ground, tens to hundreds of kilometers across that commonly contain streamlined remnants of the pre-existing terrain. Most start full size and have no tributaries. Because of their close resemblance to terrestrial flood features, they are widely, although not universally, believed to have formed by huge floods. Valley networks are comprised of valleys that are typically only 1-5 km across, although they may be hundreds even thousands of kilometers long. The networks they form resemble terrestrial river systems in plan. They are thought to have formed mostly by slow erosion of running water. Gullies are smaller still. They are restricted to steep slopes and are only several to tens of meters wide and hundreds of meters long.

### 2.2. Outflow Channels

Outflow channels vary greatly in size. They occur in several regions of the planet and start in several different types of geologic terrains (Figure 1). They clearly do not all originate in the same way. The most prominent concentration of outflow channels is around the Chryse basin. The largest outflow channel on the planet, Kasei Vallis, enters the Chryse basin from the west, and several large channels emerge from the cratered uplands to the south of the basin and affect an 800 km wide swath of terrain clearly marked by longitudinal scour and numerous teardrop shaped islands. The southern channels extend northward across the Chryse basin to 25 N where they curl northwestward and merge with the scoured ground from Kasei Vallis, and continue to roughly 40 N where traces of the channels become lost in the northern plains. Crater counts and translation relations suggest that most of the circum-Chryse channels are from the Hesperian (the middle epoch in Mars history ending sometime between 3.5 and 1.8 billion years ago). Various estimates have been made of the discharge $Q$ implied by the large size of the channels. The estimates are based on relations observed for terrestrial rivers, after correcting for the lower gravity on Mars (Komar, 1979). The basic equation is

$$Q = A \sqrt{\frac{g_m S R^{4/3}}{g_e R^2}};$$  \hspace{1cm} (1)

where $A$ is the cross-sectional area of the flow, $g_m$ and $g_e$ are gravity on Mars and Earth, $S$ is the local slope, $R$ is the hydraulic radius (ratio of cross-sectional area to wetted perimeter and $n$ is the Manning roughness coefficient (Williams et al., 2000). The roughness coefficient takes into account factors such as the roughness of the stream bed and the sinuosity of the channel and is determined empirically from terrestrial rivers. Applying this equation to Mars is uncertain because what value is appropriate for the Manning coefficient is uncertain, the depth of the floods is usually unknown and Equation 1 was determined from terrestrial rivers,
which are orders of magnitude smaller than the martian channels. For Kasei Vallis estimates for the peak discharge range from $10^4$ m$^3$ s$^{-1}$ (Williams et al., 2000) to $10^9$ m$^3$ s$^{-1}$ (Robinson and Tanaka, 1990). The total amount of water involved is even more uncertain. Roughly $6 \times 10^5$ km$^3$ was removed to form Kasei Vallis, suggesting that at least $10^6$ km$^3$ of water was needed and possibly considerably more.

Many of the valleys that extend into the Chryse basin (e.g., Maja Vallis, Sabatana Vallis, Ravi Vallis, and Ares Vallis) start full size in rubble filled hollows (Fig. 2a). The relations suggest that the floods were caused by catastrophic eruptions of groundwater followed by collapse of the source areas (Carr, 1979). To achieve the discharges needed the groundwater must have been under high pressure, possibly as a result of being trapped under a thick cryosphere. Others valleys (Tiu Vallis, Simud Vallis, and Kasei Vallis) can be traced upstream into substantial canyons. In these cases, the relations suggest catastrophic drainage of former lakes within the canyons. The presence of former lakes within Vallis Marineris (and in Juventae Chasma) is supported by the presence of extensive internal layered deposits (McCauley, 1978; Malin and Edgett, 2001) containing gypsum and kieserite (Gendrin et al., 2005). The lakes may have been ice-covered. Crater counts and superposition relations indicate that most of the outflow channels in the Chryse region are upper Hesperian in age (Scott and Tanaka, 1986), probably 2.0-3.8 billion years old (Hartmann and Neuckum, 2001).

Several outflow channels in the Amazonis-Elysium region appear to have a different origin from those around Chryse since they start at graben. The largest is Mangala Vallis (Fig. 2b). It starts abruptly at a 10 km wide notch in a graben wall then extends over 1000 km to the north, being in places up to 100 km wide. Like the Chryse channels it has a rich array of teardrop-shaped islands, convergent and divergent striations and other streamlined forms. Other smaller channels, Athabasca Vallis (10 N, 158 E) and Grjota Vallis (16 N, 163 E) also start at graben among and to the west of the hills that separate Amazonis Planitia from Elysium. The channels can be traced for hundreds of kilometers before they disappear in low-lying Cerberus plains just north of the plains-upland boundary at 170 E. From these plains emerges yet another large channel, Marte Vallis that extends north-eastward into Amazonis Planitia. These are the youngest outflow channels on the planet. Burr (2002), using the the crater chronology of Hartmann and Neuckum (2001), estimate that the ages of Athabasca Vallis, Grjota Vallis and Marte Vallis are 2-8 million years, 10-40 million years and 35-140 million years, respectively. If only approximately correct, these ages are so young as to suggest that outflow channels such as these could form today. Other channels in Tharsis, such as the Olympia Fossae and the Gordii Fossae, also start at graben. As with Athabasca Vallis, channels with streamlined forms clearly formed from fluids that erupted from the graben, but in these cases, whether the fluid was lava or water is more uncertain.
Figure 1. Left: Global map of outflow channels. Around the Chryse basin only the largest channels are named (MOLA).
Figure 2. Left: Maja Vallis and Juventae Chasma. Maja Vallis emerges from a 5 km deep rubble-filled depression. The outlet is at an elevation 4 km above the floor of the depression, so after Maja Vallis formed a lake must have been left in the depression, consistent with detection of sulfates there (Gendrin et al., 2005). The relations indicate that Maja Vallis formed by a massive eruption of groundwater (MDIM2). Right: Mangala Vallis. The channel starts at a notch in a graben, which suggests eruption of groundwater as a result of faulting and/or dike injection. (Mars Thermal Emission Imaging System, THEMIS).

Eruption of water from graben could result from several causes. Faulting may simply have disrupted the cryosphere seal over a deep aquifer, thereby allowing water from the underlying cryosphere to reach the surface. The water may have been under high pressure because of the regional topography. Alternatively, tectonic forces may have pressurized the aquifer causing the water to flood to the surface when faults broke the seal (Hanna and Phillips, 2005). Another possibility is that pressurization resulted from the injection of dikes that accompanied formation of the graben (Head et al., 2003).
Several valleys, including the Granicus, Tinjar and Hrad Valles emerge from graben on the western flank of the Elysium dome. At higher elevations, the graben are typical of those elsewhere on the planet, being narrow, steep-walled depressions. To the west, at lower elevations, however, they transition into channels with streamlined walls, teardrop-shaped islands and scoured floors. Many have broad textured rims, with lobate outer margins, as though the channels had overflowed and left a deposit on the rim. The latter observations suggested to Christiansen (1989) that the channels were cut by or at least were utilized by lahars or mudflows. Fluvial landforms are generally restricted to elevations less than -3400 m (relative to the 6.1 mbar datum), which suggests that this was the elevation of the local water table when the valleys formed (Russell and Head, 2003). As with the channels discussed in the previous section, formation of the valleys likely resulted from tectonic disruption of the cryosphere, possibly accompanied by dike injection.

The most puzzling features in this area are the Hephaestus Fossae and Hebrus Vallis (Figure 4a). The upper portions of both these valleys are fluvial in appearance but the lower portions are distinctly non-fluvial. They consist of linear segments that meet at high angles to form a pattern like a cracked pavement, except that the individual segments commonly consist of lines of unconnected depressions. The origin of the networks is unknown. Lava tubes form lines of depressions but not linear networks such as we see here. The pattern more resembles karst. They may indicate subsurface drainage through soluble rocks, such as carbonates, or through ice rich rocks.

Three large valleys, the Dao, Niger and Harmakhis Valles, start on the east rim of Hellas and extend for over 1000 km down into the floor of the Hellas basin. The Dao and Niger Valles both start near the volcano Hadriaca Patera, suggesting that the volcano was somehow connected with their origin, possibly causing local melting of ground ice or explosive release of groundwater following injection of hot magma into the hydrosphere and cryosphere.

Thus outflow channels occur widely across the planet with ages that range from over 3.5 billion years almost to the present. Most can plausibly be interpreted as the result of flooding by release of groundwater that followed tectonic, volcanic, or other disruptions of the cryosphere. High hydrostatic pressures in the underlying hydrosphere could have resulted from a variety of causes, including regional topography, tectonic deformation and volcanic intrusions. Another possible cause of flooding, particularly adjacent to the Valles Marineris, is catastrophic release of water from lakes. The lakes may have formed also by release of groundwater, possibly accompanying the massive faulting that caused much of the canyon relief. If a cryosphere was present when the lakes formed they would have quickly become ice covered. Most of the floods likely occurred under cold climatic conditions when a thick cryosphere was present. This conclusion is consistent with formation by massive eruptions of groundwater which had to be contained prior to eruption, with the presence of geologically recent outflow channels such as Athabasca, and with low erosion rates (Golombek and Bridges, 2000) and low weathering rates (Haskins et
al. 2005) of the basalts on the Hesperian plains in Gusev, which suggests that if
the surface experienced warm episodes during the Hesperian and Amazonian (the
latest epoch in Mars history), then these episodes were short. The larger channels
must have left large bodies of water at their termini, which would have rapidly
frozen, given the likely presence of a thick cryosphere.

2.3. VALLEY NETWORKS

The second class of water-worn feature is the valley network. Most of the cratered
uplands are dissected by networks of branching valleys, no more than a few kilome-
ters wide but up to hundreds, even thousands of kilometers long. Most are readily
distinguishable from the outflow channels. They occur mostly in the cratered up-
lands but are distributed unevenly. Northwest Arabia and large areas to the west
and southeast of Hellas, for example, are sparsely dissected whereas the broad
swath of terrain just south of the equator from 340 E eastward to 180 E is highly
dissected (Figure 3). Most valleys are short and drain into local lows. However,
several valleys extending down the regional slope through Terra Meridiana toward
the Chryse basin are over 1000 km long. Valleys are typically 1-4 km wide, have
cross sectional shapes that range from V-shaped in the upper reaches to U-shaped
or rectangular in the lower reaches (Figure 4b). In planimetric form they resemble
terrestrial river systems. Drainage densities range widely up to values at the low
end of the terrestrial range (Hynek and Williams, 2001). Some of the most densely
dissected surfaces are on the anks of volcanoes.

The ages of the valleys are difficult to determine. The vast majority of the val-
leys are in Noachian terrain, which is mostly dissected with some exceptions, as
noted previously. In contrast, Hesperian plains such as Syria, Solis, Hesperia and
Launae Plana are largely undissected. While these generalizations are valid, there
are numerous exceptions. Some of the largest and freshest-appearing valleys such
as Nanedi and Nirgal Valles cut Hesperian terrain. Well-developed, dense networks
cut Hesperian plains at the southern end of Echus Chasma (Mangold et al., 2004)
and on some volcanoes such as Alba Patera, Amazonian surfaces are dissected. It
appears that the dominant period of valley formation was in the Noachian and that
the rate fell dramatically at the end of the Noachian but formation continued either
episodically or at a very low rate for much of Mars history. Quantitative measures
of stream profiles and basin shapes suggest that most martian valley networks are
less well developed than their terrestrial counterparts (Stepinki and O’Hara, 2003).
In most terrestrial basins the local slope $S \sim A^{-\phi}$ where A is the area upstream at
the given point and $\phi$ is called the concavity exponent, which is a measure of how
concave the basin is. The values for martian basins (0.2-0.3) are consistently less
than terrestrial values (0.3-0.7) indicating poor basin development. Another indi-
cator of basin development is how the circularity of the basin varies with elevation.
The higher the elevation slice through a typical terrestrial basin, the more circular
the basin outline. This tendency is significantly less with martian basins.
Figure 3. Global map of valley networks. Most are in the southern uplands, although western Arabia and the uplands between Argyre and Hellas are only poorly dissected. The plains are largely undissected (MOLA).
Figure 4. Left: Utopia Channels. Hephaestus Fossa and Hebrus Valles have both fluvial and tectonic characteristics. Both start at irregular depressions, from which emerge fluvial-like channels. The Hephaestus Fossae change downstream into an array of discontinuous, intersecting linear segments. The Hebrus Valles similarly terminate in linear segments. The cause of the tectonic-like patterns are unknown. (MDIM2). Right: Warrego Valles. The terrain here is densely dissected by small valleys. Such area-filling dense dissection is characteristic of fluvial regimes in which surface runoff dominates and precipitation was widespread (THEMIS).

Although the valleys were almost certainly cut by water, the source of the water and the conditions under which formation took place remain controversial. Water could have been introduced onto the surface in three ways; as groundwater seepage, as rainfall or as snowfall. That groundwater seepage played a prominent role in development of some of the valley networks was recognized early from the open networks, the amphitheater-like terminations of tributaries (Figure 5a), rectangular cross sections, and other properties (Sharp and Malin, 1975; Pieri, 1980). Some authors have argued that the valley networks could form exclusively by groundwater sapping under conditions similar to those that prevail today (Squyres and Kasting, 1994; Gaidos and Marion, 2003). However, formation of valley networks by groundwater sapping alone seems unlikely. Dense area-filling networks (Figure 4) are common and normally do not form where seepage dominates (Craddock and Howard, 2003; Hynek and Williams, 2001). In addition, many valleys start at local highs such as crater rims and central peaks where groundwater seepage is unlikely. Moreover, seepage draws down the local water table so some form of recharge is needed to sustain erosion. Although Clifford (1987) suggested that the global hydrosphere could be recharged by basal melting of ice at the poles, his concern was mainly for post-Noachian times, not for the Noachian era when most of the valleys formed. Many of the valleys are also at higher elevations than the base of the south polar layered terrain so could not provide the hydrostatic head needed to enable seepage (Carr, 2002).

Arguments against rainfall are several: (1) It is difficult to warm early Mars when most of the valleys formed because of the faint, young Sun (Kasting, 1991);
(2) the early Mars atmosphere is vulnerable to blow-off by large impacts; (3) until recently, hydrated minerals, evaporites and carbonates had not been detected from orbit; and (4) easily weathered olivine has been detected from orbit. Craddock and Howard (2003) argue, however, that the geomorphic evidence for surface runoff and precipitation is so compelling that some assumptions in the modeling studies must be wrong. They also suggest that observational artifacts are hindering our ability to detect weathered minerals from orbit. The arguments against precipitation have been recently undermined by the finding of evaporites and water lain sediments in Meridiani (Squyres and Knoll, 2005) and by detection of evaporite minerals and phyllosilicates from orbit by the near-infrared spectrometer OMEGA on Mars Express (Bibring, 2005).

Precipitation may have been rain or snow. Snowfall can occur during the present epoch. The obliquity may have been as high as 45° within the last 10 million years. (Laskar et al., 2002) At a high obliquity, water would be driven from the poles and deposited as ice (snow) at lower latitudes. Melting of such snow could, in principle, provide meltwater to cut valleys. However, unless these changes are accompanied by significantly warmer surface conditions, a snow cover is unlikely to produce sufficient meltwater to cut valleys (Clow, 1987), although there may be sufficient water to cut gullies on steep, poleward-facing slopes as discussed next. Basal melting of snow from internal heat flow is also unlikely to produce meltwater in sufficient quantities because of the low internal heat flow. Climate change appears to be needed to form valleys. This conclusion is consistent with the patterns of dissection, the high Noachian erosion rates, detection of phyllosilicates produced by chemical weathering and finding of waterlain sediments at Meridiani.

Greenhouse warming by a CO₂-H₂O atmosphere, however, may not have been sufficient to allow rainfall (Kasting, 1991; Haberle, 1998). Other greenhouse gases may have been involved or greenhouse warming may not have been the primary mechanism for warming the planet. An alternative is that the warming resulted from the injection of large amounts of hot rock and water into the atmosphere during large impacts (Segura et al., 2002). The proposal is attractive because it is consistent with formation of the valleys mainly in the Noachian and continuing at a low rate subsequently.

2.4. GULLIES

Finally, there are the gullies. Gully is the term applied to small, linear, seemingly young erosion features incised into steep, mostly poleward-facing slopes at mid to high latitudes. They typically consist of an upper theater-shaped alcove that tapers downward to converge on one or more channels that are mostly meters to tens of meters wide and hundreds of meters long (Figure 5b). They are much smaller than the valleys just discussed and confined to steep slopes. Malin and Edgett (2000) originally ascribed them to groundwater sapping, but there is considerable uncertainty about how they formed. With mean annual temperatures close to 215 K
the ground is frozen to kilometer depths for all plausible heat flows and thermal conductivities. Moreover many gullies occur on central peaks, or reach to the lip of craters where groundwater is unlikely. They may have formed during periods of high obliquity from the summer melting of snow that accumulated on slopes in winter (Costard et al., 2002; Christensen, 2003). According to this model, they form preferentially on steep poleward facing slopes because these are constantly illuminated in summer at high obliquity.

![Figure 5](image.png)

**Figure 5.** Left: Nirgal Vallis at 27.8 S, 316.6 E. The tributaries have alcove like terminations. They do not divide into ever smaller valleys as in the previous picture. The relations suggest origin by groundwater sapping. Nirgal Vallis is Hesperian in age (MOC E0202651). Right: Gullies on the south-facing wall of Nirgal Vallis at 29.7 S, 321.4 E. The scene is 2.3 km across. The gullies are mostly incised into talus below a bedrock outcrop at the top of the slope. Fans of debris eroded by the gullies appear to be superimposed on dunes in the floor of the main valley (MOC).

Regarding ice, neutron and gamma-ray data indicate that ice is abundant at high latitudes at depths of roughly 1 meter, the detection limit from orbit. Geologic evidence suggests that ground ice is also abundant to depths of at least hundreds of meters. One indicator is a general softening of the terrain at mid to high latitudes where ice is stable (Squyres and Carr, 1986). The softening has been attributed to ice-abetted creep of near-surface materials. Another indicator of ground ice is the abundance of features at mid to high latitudes that indicate glacier-like flow of materials shed from slopes. In addition, a wide array of landforms in the northern plains, Argyre, and Hellas have been attributed to glaciers on the surface, some of which likely formed by the freezing of lakes fed by large floods (Kargel et al., 1995). Glacial features have also been identified on and adjacent to volcanoes in
Tharsis, but these had a different origin, having possibly formed by accumulation of ice during periods of high obliquity (Head et al., 2005) (Figure 6a).

2.5. LAKES AND OCEANS

What about lakes and oceans? Numerous standing bodies of water of widely different sizes must have accompanied formation of the outflow channels and valley networks. In the cratered uplands valley networks commonly converge on local lows where water likely ponded. On a larger scale lakes must have formed at the ends of outflow channels. A common relation in the uplands is the breaching of crater walls by valleys, both incoming and outgoing. In either case, the crater must have formerly contained a lake. Breaks in slope on crater walls that might be shorelines internal to craters are rare, although occasionally present. More common are deltas deposited from streams entering craters (Figure 6b) (Malin and Edgett, 2003; Hauber et al., 2005). In addition, most large upland craters have shallow flat floors. Where the floors are eroded, they are commonly revealed to be underlain by finely layered deposits. While not proof of lacustrine deposition, they are consistent with it. In some areas, such as around Meridiani, mounds of sediments are common within craters. They appear to be remnants of former regional deposits such as were sampled by “Opportunity”, and may contain lacustrine deposits as do those at the Meridiani landing site.

Figure 6. Left: Possible glacier east of Hellas at 38 S, 104 E. Material has flowed from an alcove in a massif in the upper right into a nearby crater, and from there through a gap into another larger crater 500 m lower in elevation. Right: Former crater lake at 8.5 N, 312 E. A delta formed within the crater from material brought down the channel that enters through southern rim. Such must have been common throughout the uplands in the Noachian when most of the valley networks formed (THEMIS).
The former presence of much larger bodies of water has been suggested for the northern plains (Parker et al., 1989; Parker et al., 1993; Clifford and Parker, 2001), for Hellas (Moore and Wilhelms, 2001) and Argyre (Parker et al., 2000). Their former presence is not controversial, although their sizes are. Estimates for the size of the northern ocean range from roughly $2 \times 10^7$ km$^3$ for the Deuteronilus shoreline of Clifford and Parker (2001) (Figure 7) to $3 \times 10^8$ km$^3$, based on the assumption that the Olympus Mons cliff was wave cut (Baker, 2001). The Deuteronilus shoreline, as mapped by breaks in slope and textural changes encloses an area where deposits of upper Hesperian age, interpreted as effluent from large floods, have partly buried craters and ridges (Head et al., 2002). Thus the shoreline mapped from breaks in slope and similar features is supported by the regional geology. It would have enclosed roughly $2 \times 10^7$ km$^3$ of water. Evidence for larger bodies of water is primarily based on alignment of terraces, benches, escarpments and so forth. Their validity as shorelines is difficult to assess because of the multiple ways that such features can form. Moreover, bodies of water larger than that outlined by the Deuteronilus shoreline are not required by the dimensions of the channels entering the basin.

As indicated previously, the large outflow channels around the Chryse basin are mostly upper Hesperian in age as are the sediments identified as effluent by Head et al. (2002). Most of the shoreline features identified by Parker and his co-workers are also post-Noachian. Thus a plausible case can be made for large bodies of water in the northern plains in the upper Hesperian. Evidence for earlier oceans in the Noachian, for which there is stronger evidence for warm conditions, is much weaker. Yet if warm conditions prevailed, as strongly suggested by the valley networks, then large bodies of water must have been present in low areas such as the northern plains. The evidence has probably been largely erased by subsequent events. The proposed shorelines in Hellas are more continuous than those around the northern basin. Moore and Wilhelms (2001) recognized two strandlines, one at -5800 m and one at -3100 m. Both can readily be traced at a constant elevation around a substantial fraction of the basin.

The presence of young outflow channels such as Athabasca, Grjota and Marte Valles implies that lakes were present at least transiently, and ice-covered in the recent geologic past. Interpretation of some platey features as ice flows (Murray et al., 2005) where such lakes may have been present is, however, controversial.

2.6. Surface Observations

We now have observations from the surface that pertain to the hydrologic history of Mars. The findings of the rover “Opportunity” in Meridiani strongly support the presence of liquid water at the surface in the late Noachian. The rover landed on a sequence of rocks that unconformably overlie cratered and dissected Noachian rocks. Crater counts suggest that the sequence dates back to close to the Noachian-Hesperian boundary, or roughly 3.5-3.7 billion years old. While most of the se-
Figure 7. Possible shoreline. This view, looking down on the north poles, shows the -3760 m contour, which roughly corresponds to the lowest shoreline proposed by Parker et al., 1989, 1993). The volume below the contour is $2 \times 10^7$ km$^3$, and is indicative of the size of a body of water that would be left after one of the largest floods.

The findings of the rover “Spirit” at Gusev are more ambiguous. The rocks on the Hesperian age floor of Gusev are basalts that are almost pristine except for a
millimeter thick, sulfate-rich rind. The rocks of the Columbia Hills are, however, very different. They vary greatly in type from dunites to sulfate-rich rocks, and their silicic components range from pristine primary minerals, to highly oxidized and hydrated minerals. Many of these rocks have clearly been altered under warm aqueous conditions, but whether these conditions occurred at the surface or deep below the surface is uncertain in most cases, as are the climatic conditions, if any, that are implied (Squyres et al., 2006).

Finally, there is evidence based on erosion rates. Evidence for a dramatic change in erosion rates at the end of the Noachian is unambiguous. On Hesperian and Amazonian terrains most impact craters are almost perfectly preserved, including the most subtle textures on their ejecta blankets. In contrast, in Noachian terrains almost all the larger older craters are highly eroded, many being detectable only by subtle topographic signatures (Schultz and Frey, 1990). Golombek and Bridges (2000) estimate that the erosion rates decreased by 3-6 orders of magnitude at the end of the Noachian.

2.7. Hydrologic and Climatic Change on Mars

To summarize what we know about hydrologic history: That a dramatic change in the hydrologic regime occurred at the end of the Noachian cannot be doubted. The rate of valley formation declined dramatically, erosion rates fell by orders of magnitude, and weathering rates as indicated by production of phyllosilicates also fell. Widespread, dense dissection of the Noachian terrains, implies an active Noachian hydrologic cycle with precipitation, surface runoff, infiltration, movement of groundwater and accumulation of standing bodies of water, all consistent with warm climatic conditions. In contrast, the scarcity of valley networks, low erosion rates, lack of weathering products, and presence of outflow channels that appear to have formed mostly by massive eruptions of groundwater, all suggest that most post-Noachian times were characterized by cold surface conditions, a thick cryosphere, and the occasional large floods that created bodies of water that rapidly froze.

Despite the likely warm conditions and an active hydrologic cycle during the Noachian, large integrated drainage basins comparable to those of the Mississippi, Amazon, and Nile did not develop. Most of the valleys in the Noachian terrain drain into local lows. They typically do not merge with other networks to form large basins thousands of kilometers across, as has happened in many areas on Earth. The immaturity of the drainage system is also reflected in the poor concavity and relations between circularity and elevation within the basins as described previously. These indications of immaturity may simply reflect the balance between the rate of terrain creation, which in the Noachian was largely by impacts and volcanism, and the rate of terrain degradation by fluvial erosion. On Noachian Mars the balance appears to have been more in favor of terrain creation than on present-day Earth.
Whether this was the result a higher rate of terrain formation or a lower rate of terrain degradation is unclear.

The extent of any Noachian oceans is also unclear. However, if there was an active hydrologic cycle as appears likely, then there must have been large bodies of water in low areas. Much of the near-surface inventory of water on Mars today is likely sequestered as ice in the kilometers thick cryosphere or trapped in the hydrosphere below (Clifford, 1993). If the surface of Mars was warm and wet in the Noachian then much of the water now locked in or under the cryosphere would have been able to percolate into the low-lying basins such as Hellas and the northern plains and so participate in the global hydrologic cycle. The Meridiani shoreline around the northern basin, proposed by Clifford and Parker (2001), if real, would be Noachian in age. It encloses the equivalent of a global layer 1.5 km thick. If this was truly a Noachian shoreline, then approximately half the planet would have been covered with water. If the upper strandline in Hellas was close to the global sea level then the volume of water would have been 0.5 km spread evenly over the whole planet, and 20% of the planet would have been under water.

In addition to large bodies of water in basins such as Hellas and Isidis, there would have been numerous small bodies of water in the hollows of the poorly graded landscape. The presence of lakes within many Noachian craters is indicated by valleys entering or leaving the craters. Many of these local lows, including craters, now contain finely layered deposits of yet to be determined origin.

The Noachian rocks of the Columbia Hills in Gusev Crater are interbedded volcanic rocks and impact breccias that show a wide range of aqueous alteration, as would be expected from an era of high rates of volcanism and impacts and abundant near-surface water. High rates of hydrothermal activity are also expected although no unambiguous hydrothermal deposits have been detected.

At the end of the Noachian conditions changed dramatically. The most characteristic hydrologic feature of the Hesperian is the outflow channel. Valley formation became highly localized, sapping characteristics became more common, erosion rates fell dramatically, and phyllosilicates production declined although sulfates deposits continued to form. The change from a regime dominated by surface runoff to one dominated by outflow channels, many of which appear to be eruptions of groundwater, could be explained by a change in the global climate at the end of the Noachian and development of a thick cryosphere. However, while such a change likely did occur, the story is more complicated. There are local, highly dissected post-Noachian surfaces. Many of the most densely dissected on the planet are on post-Noachian volcanoes such as Ceraunius Tholus, and Hecates Tholus. In addition, there are rare, local, highly dissected surfaces away from volcanoes, such as the Hesperian plains adjacent to the southern end of Echus Chasma (Mangold et al., 2004).

Most of the fluvial features of the post-Noachian era are consistent with the presence of a thick cryosphere. Eruption of water from below the cryosphere as a result of tectonic activity, volcanism, or impacts may have been catastrophic as in
the case of the large outflow channels, or more gentle as with valleys with sapping characteristics such as Nirgal Vallis and Nanedi Vallis. Outflow channels continued to form throughout much of Mars’ history, although the younger outflow channels such as the Athabasca and Gordi Valles are much smaller than those that formed in the late Hesperian. Water brought to the surface would have pooled wherever it encountered hollows, either at the start of the channels (Juventae Chasma) or at the ends of the channels (the northern plains). Finding of sulfates within the canyons and other hollows is consistent with sublimation of lakes formed by groundwater brought to the surface. Lack of hydrous silicate minerals in post-Noachian terrains is consistent with a dominantly cold climate and a thick cryosphere.

While this simple picture explains much of what we see, there are anomalies as previously noted. Various proposals have been made to explain young valley systems: (1) The young valleys on volcanoes result from circulation of groundwater onto the surface as a result of hydrothermal activity (Gulick, 1998). (2) Formation of large outflow channels by CO₂-charged floods episodically caused temporary changes in the global climate (Baker, 2001). (3) The global climate was episodically and temporarily changed by large impacts (Segura et al., 2002) or massive volcanic eruptions. (4) The valleys formed by melting of ice during periods of high obliquity (Jakosky and Carr, 1985). There may be other possibilities. However, despite these possible excursions, for most of the Hesperian and Amazonian, Mars appears to have been cold, with a thick cryosphere, and extremely low rates of erosion and weathering.

Guillame: Thank you, William. Now we have Charles Lyell. Sir Lyell received both his B.A. and M.A. from Exeter College. After a brief legal career he devoted himself to geology and quickly became an established proponent of what was later known as the uniformitarianism view in which geologic phenomena were explained by the cumulative effect of processes acting gradually over an immense period of time. He was elected a fellow of the Royal Society in 1826 and began publishing his well-known Principles of Geology in 1830. Charles?

3. Volatile Reservoirs and Exchange Processes on Earth and Mars

Lyell: Cheers. Geological processes that sustain reservoirs of volatiles and their geochemical cycling in planets are part of the “life support system” that can make planets habitable. To sustain life a planet must provide key chemical constituents as well as environmental conditions and sources of energy that can fuel biological processes. Igneous rocks that are abundant in the crusts of silicate-rich planets in our inner Solar System offer many of the needed elemental ingredients. Silicate rocks provide phosphorous and essential metals such as iron, magnesium, calcium and key trace elements. However organisms consist principally of water and compounds of the elements C, N, S, and P, and silicate rocks typically have
relatively minor amounts of these. Although the Earth’s atmosphere, hydrosphere and sedimentary rocks contain abundant water, oxidized C compounds, N, Cl, and various S compounds, these substances clearly did not derive principally from crustal igneous rocks. W. Rubey proposed that such “excess volatiles”, including water, derived from volcanic outgassing and other thermal processes that have been sustained by Earth’s geologic activity (Rubey, 1951).

These “excess volatiles” are significant also because they helped to create and maintain habitable environmental conditions at and near Earth’s surface. They warm and moderate Earth’s climate and thereby create temperatures and pressures necessary to stabilize liquid water, one of the key ingredients essential for life as we know it. These volatiles also interact with the solid planet in ways that can buffer environmental perturbations caused by volcanism, large impacts and long-term changes in solar luminosity.

Habitable planetary environments also must provide sources of energy that living systems can utilize to drive their metabolism. Today’s biosphere is dominated by photosynthetic biota that can harvest abundant solar energy. However many other organisms are non-photosynthetic and can obtain useful chemical energy by reacting oxidized and reduced chemical compounds in so-called “redox” reactions. Processes of volatile exchange in the Earth’s crust have delivered both oxidized and reduced chemical compounds to habitable environments.

A proper understanding of the roles played by volatile exchange processes requires that they be considered in the context of a network of volatile reservoirs in the atmosphere, hydrosphere, crust and mantle that are linked by physical and chemical processes. Such networks are called geochemical cycles because a key dynamic is that volatile constituents can move back and forth between two or more reservoirs over time. On Earth such networks are called biogeochemical cycles because life itself is an important crustal process. Global biogeochemical cycles actually consist of multiple nested cyclic pathways that differ with respect to their reservoirs and processes. However, all pathways ultimately pass through reservoirs of volatiles in the hydrosphere and atmosphere, and these shared reservoirs unite all of the sub-cycles and allow even their most remote constituents to influence the biosphere.

3.1. BIOGEOCHEMICAL CYCLING OF VOLATILES ON EARTH

The present-day biogeochemical cycles of volatiles on the Earth are represented graphically in Figure 8 as an integrated system of reservoirs and processes. Reservoirs and processes are shown as boxes and labeled arrows, respectively, that delineate the various sub-cycles, denoted in the figure by the labels “HAB”, “SED”, “MET” and “MAN”. The range of timescales typically needed for a particular chemical constituent to traverse each of these sub-cycles is indicated along the right margin, below their corresponding labels. Of course the actual physical boundaries between these sub-cycles are not so sharply delineated in nature. But the sub-cycles
depicted here do represent characteristic domains along the continuum of reservoirs and processes that collectively constitute the Earth system.

Figure 8. Schematic diagram of Earth's biogeochemical cycles of volatiles, showing reservoirs (boxes) in the mantle, crust, oceans and atmosphere, and showing the processes (arrows) that unite these reservoirs. The vertical bars at right indicate the timeframes within which a volatile element or compound typically completely traverses each of the four sub-cycles (the HAB, SED, MET and MAN sub-cycles, see text). For example, C can traverse the hydrosphere-atmosphere-biosphere (HAB) sub-cycle typically in the time scale between 0 to 1000 years.

Now I will discuss each individual “sub-cycle”: The first includes the hydrosphere, atmosphere, and biosphere (HAB). The HAB sub-cycle includes only the volatile reservoirs in the hydrosphere, atmosphere and biosphere (Figure 8) and is characterized by typically geologically high rates of exchange between these reservoirs. Volatile constituents can traverse the HAB sub-cycle within time scales of minutes to hours, for example, during the rapid cycling of constituents between biota and their environment or across the air-sea interface. Alternatively, volatiles in seawater can require a thousand years or more to be cycled due to the circulation rate of the global ocean (Broecker and Peng, 1982). The ocean indeed exerts a dominant role in the HAB sub-cycle due to its great size and enormous stored content of thermal energy from the Sun.

The next sub-cycle in terms of timescale is the sedimentary (SED) sub-cycle which includes the entire HAB sub-cycle plus reservoirs of volatile constituents in sedimentary rocks (Figure 8). The SED sub-cycle strongly influences the HAB sub-cycle. For example, sedimentation limits global productivity by removing nutrients, phosphorus in particular. The balance between sedimentation of oxidized
(carbonate, sulfate, Fe$^{3+}$) versus reduced (organic C, sulfides, Fe$^{2+}$) species determines the abundances of O$_2$ and sulfate in the atmosphere and hydrosphere (Holland, 1984). The weathering and transport of igneous and sedimentary rocks deliver nutrients to the oceans, thereby modulating global biological productivity. Organisms can substantially enhance weathering rates. The rates of weathering and erosion of sediments and rocks, together with the contents of reduced species in those deposits, determine their rates of consumption of O$_2$ and other chemically reactive constituents.

The rate at which an atom or chemical constituent traverses the SED sub-cycle is determined principally by tectonic controls upon rates of formation and destruction of sedimentary rocks. Reduced C, N and S compounds are buried principally in deltaic-shelf sediments and in sediments beneath highly productive open-ocean regions. The global net burial rates of C and S species are controlled ultimately by the rates of delivery of those species to the oceans by weathering and transport and by global rates of their reaction with submarine basalts (Garrels and Perry, 1974). Sedimentary reservoirs of C and S are much larger than C and S reservoirs in the HAB sub-cycle (Des Marais, 2001). The average sedimentary rock survives for about 200 million years (Derry et al., 1992). Therefore key biogeochemical properties of the oceans and atmosphere, including their nutrient inventories and oxidation states, are modulated by the SED sub-cycle over timescales between tens of thousands to hundreds of millions of years (Figure 8).

Next is the metamorphic (MET) sub-cycle (Figure 8), which includes more deeply buried sedimentary and igneous rocks that are altered (metamorphosed) by elevated temperatures and/or pressures. Volatiles in the MET sub-cycle include those in rocks that enter subduction zones but ultimately escape injection into the mantle either because they are degassed or because their host rocks also escape subduction, for example, by lateral accretion into continental crust. The mass of rocks in the MET sub-cycle greatly exceeds those in the SED sub-cycle (Lowe, 1992). However inventories of reduced C in the MET sub-cycle are smaller (Hunt, 1972), due both to losses during metamorphism of sedimentary rocks and to the typically much lower reduced C contents of crustal igneous rocks. Carbonate C is also lost during thermal metamorphism as CO$_2$. Volatile constituents require typically tens of millions to billions of years to traverse the MET sub-cycle. These cycle times are typically longer than those for the SED sub-cycle and reflect the longer lifetimes of more deeply buried continental rocks.

Finally, there is the deepest mantle-crust or MAN sub-cycle which includes the mantle volatile reservoirs (Figure 8) and the processes of subduction and mantle outgassing. The modern global rates of outgassing of key volatile species from the mantle are summarized in Table I. An atom of C requires between tens of millions of years to as long as billions of years to traverse the MAN sub-cycle (Des Marais, 2001). Over timescales of tens of millions to billions of years, the processes of mantle-crust exchange probably modulated both the sizes and the overall oxidation states of the much smaller C reservoirs in the HAB, SED and MET sub-cycles.
In addition, thermal emanations of other reduced species (principally sulfides, \( \text{H}_2 \), and \( \text{Fe}^{2+} \)) consume \( \text{O}_2 \) and also contribute reducing power for biosynthesis by microbes that derive energy from redox reactions. Thus the balance between biological productivity and decomposition, sediment cycling, and thermal processes has modulated the overall oxidation state of the surface environment and the crustal reservoirs of key redox-sensitive elements such as C, N, S and Fe.

Biota that dwell at or below the sea floor along the mid-ocean ridges obtain their energy principally from oxidation-reduction reactions involving reduced hydrothermal emanations (Jannasch and Wirsen, 1979). They utilize reduced S, \( \text{H}_2 \) (derived from water-rock reactions), and \( \text{Fe}^{2+} \). Today this total flux of reduced constituents, expressed as \( \text{O}_2 \) equivalents, is in the range \( 0.2 - 2.1 \times 10^{12} \text{ mol yr}^{-1} \) (Table I) (Elderfield and Schultz, 1996).

### Table I
Mid-ocean ridge hydrothermal fluxes of reduced species to the ocean and atmosphere. Modern fluxes are from Elderfield and Schultz, (1996). aLinear interpolation between fluxes estimated for today and for 3.0 Ga. bAssumes that thermal fluxes of reduced chemical species have scaled linearly to mid-ocean ridge spreading rates, and that these spreading rates have scaled with the square of the heat flow (Sleep, 1979). At 3.0 Ga, heat flow is estimated to have been approximately 2.2 times the modern value (Turcotte, 1980).

<table>
<thead>
<tr>
<th>Species</th>
<th>Flux ( [10^{12} \text{ mol yr}^{-1}] )</th>
<th>( \text{O}_2 ) consumed ( [10^{12} \text{ mol yr}^{-1}] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{S}_{\text{reduced}} )</td>
<td>0.085-0.96</td>
<td>0.18-1.92</td>
</tr>
<tr>
<td>( \text{Fe}^{2+} )</td>
<td>0.023-0.19</td>
<td>0.02-0.19</td>
</tr>
<tr>
<td>( \text{Mn}^{2+} )</td>
<td>0.011-0.034</td>
<td>0.01-0.034</td>
</tr>
<tr>
<td>( \text{H}_2 )</td>
<td>0.003-0.015</td>
<td>0.002-0.008</td>
</tr>
<tr>
<td>( \text{CH}_4 )</td>
<td>0.007-0.024</td>
<td>0.014-0.048</td>
</tr>
<tr>
<td>Modern total</td>
<td></td>
<td>0.2-2.1</td>
</tr>
<tr>
<td>Ancient total (2.1 Ga(^a))</td>
<td></td>
<td>0.74-7.7</td>
</tr>
<tr>
<td>(2.2 Ga(^a))</td>
<td></td>
<td>0.77-8.0</td>
</tr>
<tr>
<td>(3.0 Ga(^a))</td>
<td></td>
<td>0.97-10</td>
</tr>
</tbody>
</table>

### 3.2. Changes in Biogeochemical Cycling over Time

Several processes have changed rates of volatile exchange during early Earth history. One such process is the Sun itself. Most current models of stellar evolution predict that the Sun was less luminous when it first entered the main sequence...
about 4.6 billion years ago. As the Sun burned H to He, its core became denser and therefore hotter. This increased the rate of thermonuclear burning, which, in turn, increased the Sun’s luminosity over time. According to one estimate (Gough, 1981), the sun was only 70% as luminous 4.6 billion years ago as it is today. This “faint early sun” presents a paradox for Earth’s early climate (Sagan and Mullen, 1972) because, if one assumes that parameters such as atmospheric composition and planetary albedo had been the same as today, then Earth’s mean surface temperature would have been below freezing during the first 2 to 2.5 billion years of its history. However, very ancient (ca. 3.7 billion-year-old) metasedimentary rocks indicate that large standing bodies of water were abundant (e.g., Schopf, 1983).

Another is impacts. Large impacts should have released substantial quantities of volatiles (Sleep and Zahnle, 2001). Impacts on Mars have been credited with releasing crustal volatiles (CO$_2$, H$_2$O, etc.), which helped to enhance an ancient greenhouse (Newsom, 1980; Segura et al., 2002). However impacts also created dust clouds that, in the aftermath of the impacts, might have weathered quickly, thus rapidly removing atmospheric CO$_2$ and perhaps triggering periodic profound cooling of the Hadean climate (Sleep and Zahnle, 2001). More recent impacts probably affected surface volatiles, at least over shorter timescales. For example, the impact at the Cretaceous-Paleogene boundary probably released large quantities of C and S from sedimentary carbonates and sulfates and upper mantle materials that were within the Yucatan target zone (Toon, 1997). Impacts of this size very likely affected the HAB subcycle for as long as thousands of years. Such impacts might have exerted more profound and permanent effects upon the surface environment, but interpretations of the specific details of these effects have been considerably more controversial.

A third is mantle-outgassing. Over timescales of tens of millions to billions of years, processes that govern the exchange of volatile species between Earth’s surface, deep crust and upper mantle probably affected volatile inventories in the crust, oceans and atmosphere. Mid-ocean ridge volcanism is quantitatively the most important source of mantle volatiles. The heat flow from Earth’s interior was substantially greater during the earlier Precambrian (e.g., Lambert, 1976). During the past 3.0 billion years, decay of the radioisotopes $^{238}$U, $^{235}$U, $^{232}$Th and $^{40}$K has been the principal source of this heat, therefore their decay over time has caused global heat flow to decline. Thermal fluxes of volatiles, including CO$_2$, can be scaled linearly to mid-ocean ridge spreading rates, and these rates vary with the square of heat flow (Sleep, 1979). Thus, for example, heat flow 3.0 billion years ago has been estimated to have been 2.2 times its modern value (Turcotte, 1980), therefore the mid-ocean ridge mantle CO$_2$ flux was perhaps approximately 5 times its present-day value.

A fourth process is subduction of the crust at plate margins returning volatiles to the mantle. For example, if the evolution of temperature and pressure regimes of subducting slabs of modern oceanic crust are considered together with the stability of C species in such slabs, then a substantial fraction of the subducted C could
escape dissociation and melting and be carried to considerable depths, possibly to be retained in the mantle for very long periods of time (Huang et al., 1980). However, three billion years ago, subducted C would, at any particular pressure, have experienced considerably greater temperatures (McCulloch, 1993). These greater temperatures enhanced the likelihood that subducted C reacted to form mobile phases that migrated upward and therefore escaped injection into the mantle (Des Marais, 1985).

A fifth is metamorphism. If the total C inventory in the HAB sub-cycle has been maintained over time at near-steady state, then net losses of C from the HAB sub-cycle during sedimentation and burial must have been balanced by the release of CO$_2$ during thermal metamorphism of sediments (Berner et al., 1983). For example, because carbonates decompose to CO$_2$ during the subduction of sediments, rates of CO$_2$ outgassing should vary with global mean spreading rates (which correlate with global mean subduction rates). Also, greater outgassing of CO$_2$ from deep continental interiors should correlate with greater worldwide tectonism that should correlate with faster mean global spreading rates (Berner et al., 1983). Therefore the mean global rate of CO$_2$ release from rock metamorphism should be proportional to global heat flow. Consequently, due to the long-term decline in global heat flow, the rates of transfer of CO$_2$ by thermal processes from the crustal sedimentary rocks (SED sub-cycle) to reservoirs in the HAB sub-cycle should have declined.

What are the roles played by continents? Continents are important because; (1) subaerial weathering is a key sink for volatiles; (2) rivers strongly affect seawater chemistry; and (3) continents have been much more stable repositories of sedimentary volatile reservoirs than have ocean basins. Accordingly any substantial long-term changes in the total area, thickness, stability, and subaerial exposure of continents would have contributed to major long-term changes in the cycling of volatile elements.

The large continents as we know them today required a considerable interval of geologic time during early Earth history to grow and become stabilized (thicker and less dense) by anatexis (partial melting and chemical fractionation), metamorphism, and under-plating (Lowman, 1989). Due to higher heat flow on early Earth, virtually all of the ancestral pre-stabilized crust was destroyed, largely by recycling into the mantle. These factors probably shortened the lifetimes of sedimentary reservoirs of volatiles. Higher heat flow led to higher sea floor spreading rates (Sleep, 1979) and lower mean ages of oceanic crust. Therefore typical Archean oceanic crust was younger, hotter and thus more buoyant than typical modern seafloor. An overall greater buoyancy of oceanic crust on early Earth created shallower ocean basins that, in turn, probably displaced more seawater onto the continents (Hays and Pitman, 1973). Thus even if the mass of ancient continental crust had been similar to that of modern crust, the land area on the young Earth probably was less extensive.
Volatile exchange processes have co-evolved with the global environment. The Archean rock record supports the view that the basic architecture of biogeochemical cycles, namely the nested HAB, SED, MET and MAN cycles (Figure 8) was in place before 3.5 billion years ago. The major changes over time occurred principally in the relative sizes of the various reservoirs and the fluxes between them. For example, the most direct solution to the “faint early sun problem” (Sagan and Mullen, 1972) is to invoke a stronger greenhouse effect in the early atmosphere that was sustained by substantially higher CO$_2$ levels. One-dimensional climate models have been used to estimate the CO$_2$ levels required (e.g., Kasting, 1987). For example, if the global mean temperatures 4.5 and 2.5 billion years ago were equal to the modern global mean temperature, pure CO$_2$ atmospheres of 1 and 0.1 bar, respectively, would have been required. Therefore, CO$_2$ declined perhaps by a factor of 1000 or more, from 4.5 billion years ago to the present. This large CO$_2$ decline required that the HAB and SED sub-cycles changed over time. These changes could have occurred as an expression of a self-regulating climate control system (Walker et al., 1981). For example, low solar luminosity would have favored low global temperatures that, in turn, would have reduced rates of water evaporation, precipitation and therefore lowered rates of chemical erosion of silicate rocks. Low erosion rates lowered the rate of CO$_2$ removal from the atmosphere, which would have allowed thermal CO$_2$ sources to increase atmospheric CO$_2$ levels. This would have raised surface temperatures until the rate of CO$_2$ removal by weathering achieved a balance with the thermal sources. As solar luminosity increased slowly over time, the CO$_2$ levels needed to maintain the temperature at which erosional (CO$_2$ sink) and thermal (CO$_2$ source) processes balanced would have slowly declined. If the pH of the global ocean remained reasonably constant over time, then seawater HCO$_3^-$ and CO$_3^{2-}$ levels would have declined in close parallel with the decline in atmospheric CO$_2$ levels. Such a substantial decline in seawater HCO$_3^-$ concentrations (and, therefore, CO$_3^{2-}$ concentrations) undoubtedly affected processes of carbonate precipitation (Grotzinger and Kasting, 1993).

Recently, atmospheric CH$_4$ has gained favor as potentially a very significant source of greenhouse warming during the Archean and early Proterozoic Eons (e.g., Pavlov et al., 2001a). The production and accumulation of abundant CH$_4$ would have been favored by hydrothermal activity, which was greater during the Archean and that produced CH$_4$ and H$_2$, and by methane-producing microbes that occupied widespread anoxic environments and utilized H$_2$ and other sources of reducing power.

The hotter early mantle must have influenced significantly the inventories of volatiles in the crust, oceans and atmosphere. Higher rates of crustal production were accompanied by higher rates of mantle outgassing (e.g., Des Marais, 2001). A hotter mantle retained subducted volatiles with greater difficulty (McCulloch, 1993). These considerations are consistent with an early Earth in which the crustal C inventory might even have exceeded the modern inventory (Des Marais, 1985;
Zhang and Zindler, 1993), and the cycling of volatiles between the mantle and crust was more vigorous than today.

Well-preserved sedimentary rocks are not abundant within the relatively few provinces of lithosphere older than 2.7 billion years. The best-preserved Archean sediments occur in the 3.5 to 3.2 billion-year-old Kapvaal Craton of South Africa and the Pilbara Block of Western Australia (Lowe, 1992). These deposits are associated with episodes of greenstone activity and intrusive events that created stable microcontinents or cratons. These cratons later became the nuclei of full-sized modern continents. Most of the Archean continental crust had yet to become stabilized by cratonization (Rogers, 1996), therefore a greater fraction of continental sediments experienced relatively higher rates of instability and thermal alteration. The tectonically more active regime within the Archean marine basins favored rapid destruction by continental collisions, partial melting and mantle/crust exchange (Windley, 1984).

Chemical weathering was very effective during the Archean, consistent with high CO$_2$ levels (Walker, 1985) and a warm climate (Lowe, 1992). Weathering of a typical uplifted rock sequence produced coarse clastic sediments that became enriched in the most chemically resistant components such as cherts and silicified komatiitic and dacitic tuffs (Nocita and Lowe, 1990). Despite the rapid uplift and transport of these rocks and their debris, their less chemically resistant components were efficiently degraded. The apparently highly effective weathering is consistent both with relatively warm, moist conditions and with elevated atmospheric CO$_2$ levels (Lowe, 1994). Altered evaporites also occur in greenstone sequences between 3.5 and 3.2 billion years ago (e.g., Buick and Dunlop, 1990). Their formation in such tectonically unstable environments is consistent with high rates of evaporation.

In the late Archean and early Proterozoic, following the inevitable decay of radioactive nuclides in the mantle, the heat flow from Earth’s interior declined (Turcotte, 1980). This decreased the rates of both sea floor hydrothermal circulation and volcanic outgassing of reduced species. The style of subduction also changed (McCulloch, 1993). In the early- to mid-Archean, subducted slabs were dehydrated, sustained partial melting, and largely disaggregated in the upper 200 km of the mantle. Later, the reduced heat flow and lower temperatures permitted colder, stronger oceanic lithosphere to develop. Subducting slabs thus sustained perhaps only partial dehydration and, together with volatiles such as CO$_2$ and H$_2$O, penetrated to depths exceeding 600 km (McCulloch, 1993). It has been proposed (Kasting et al., 1993) that the upper mantle was oxidized by the subduction of water, followed by the escape of reduced gases. A progressive oxidation of the upper mantle has not yet been demonstrated (Delano, 2001), but, if it had occurred, its effect upon the redox balance of volatiles would have been substantial.

The reworking of Archean continental crust by tectonism, igneous activity and metamorphism also had important consequences for the processing of volatiles. Marine carbonates of this age record a substantial increase in $^{87}$Sr/$^{86}$Sr values, in-
indicating greater continental erosion and runoff (Mirota and Veizer, 1994). New and extensive stable shallow water platforms became sites for the deposition and long-term preservation of carbonates (Grotzinger, 1989) and organic C (Des Marais, 1994). Increased global continental erosion rates also would have accelerated the rate of decline of atmospheric CO$_2$ (Walker, 1990). Increased subaerial weathering would have enhanced the delivery of nutrients to coastal waters, enhancing biological productivity (Betts and Holland, 1991). Greater productivity would have removed more CO$_2$ from surface seawater, but, given the still-higher-than-present oceanic and atmospheric inorganic C contents, the effect of this productivity on the atmosphere should have been minor.

Patterns of carbonate deposition, as well as the presence or absence of gypsum and anhydrite in associated evaporites, indicate that seawater concentrations of HCO$_3^-$ and CO$_3^{2-}$ have declined and SO$_4^{2-}$ has increased since the late Archean (Grotzinger and Kasting, 1993). Late Archean platform sequences include relatively abundant evidence of abiotic carbonate precipitation as tidal flat tufas and marine cements. Evaporite sequences often proceed directly from carbonate to halite deposition, thus excluding gypsum/anhydrite deposition. These observations are consistent with significantly lower seawater SO$_4^{2-}$ concentrations and/or considerably greater HCO$_3^-$ concentrations. In either case, the ratio of HCO$_3^-$ to Ca$^{2+}$ was sufficiently large to prevent deposition of gypsum/anhydrite in marine or marginal-marine environments. These observations are consistent with the view that inorganic C reservoirs within the HAB subcycle were much higher during the Archean and Paleoproterozoic (Walker 1985).

During most of the Archean Eon (prior to ca. 2.7 billion years ago), SO$_4^{2-}$ was locally present in shallow seawater at concentrations well below 1 mM (Canfield et al., 2000), considerably less than modern seawater concentrations (~27 mM). Locally high SO$_4^{2-}$ concentrations were precipitated as evaporitic sulfate minerals. The deposition of gypsum rather than anhydrite (Lowe, 1983) indicates that temperatures very likely were below 58$^\circ$C. Sulfate concentrations increased between 2.5 and 2.1 billion years ago, perhaps in response to extensive oxidation of sulfur due to widespread oxygen-producing photosynthetic biota. Molecular oxygen emanating from these communities might have hastened the rate of destruction of atmospheric CH$_4$, leading to global cooling and the onset of glaciation about 2.5 billion years ago (Kasting and Catling, 2003). Still, the deposition of banded iron chemical sediments persisted until ca 1.8 billion years ago and indicates that the deep oceans prior to that time contained appreciable concentrations of dissolved Fe$^{2+}$. After 1.8 billion years ago, levels of seawater SO$_4^{2-}$ were sufficient to enable sulfate-reducing microbes to produce sulfide at rates that exceeded the rate of supply of Fe$^{2+}$ to the global oceans (Canfield and Raiswell, 1999). Thereafter, banded iron formations disappeared and the deep oceans remained anoxic and sulfide-rich until perhaps sometime after 800 million years ago (Canfield and Raiswell, 1999). Thereafter, a series of global climate perturbations ensued and included episodic widespread glaciations. These ended just prior to the dawn of the Phanerozoic Eon.
(543 million years ago), during which the deep ocean became more oxidized as it approached its modern state.

3.3. The Geochemistry of Volatiles on Mars

That is our current picture of Earth volatiles. I’ll now discuss cycling of volatiles on Mars. Mars is perhaps the most likely planet to provide a second example in our solar system of a habitable planet. Mars is the only major planet other than Earth where evidence indicates that liquid water contributed substantially to the development of its crust. But there are important differences. For example, today, Mars clearly has no ocean or smaller standing bodies of water. Because Mars is smaller and has lower a gravitational field than Earth, and because it lacks an active magnetic dynamo, losses of volatiles to space by various mechanisms (Jakosky and Jones, 1994) have been far more extensive than on Earth. Mars’ smaller size and greater distance from the Sun can help us begin to understand how processes that exchange volatiles have influenced the habitability of diverse planets. It is important to determine whether the exchange of volatiles on Mars could ever have provided the key chemical constituents, environmental conditions and sources of energy that could have sustained life.

Volatiles in the Martian atmosphere and surface have exchanged with reservoirs elsewhere in the planet, some volatiles have reacted irreversibly with crustal materials, and some have been lost to space permanently (Figure 9). Martian meteorites and spacecraft observations have provided clues about reservoirs of volatiles in the atmosphere and crust as well as the processes that linked them. Hydrothermal systems sustained by volcanism should have been widespread on Mars (Gulick, 1998). Volcanic aerosols have long been proposed as the source of the high S and Cl (“excess volatiles”) in the Martian soil (Clark and Baird, 1979) and the abundant sulfate found at both of the Mars Exploration Rover landing sites. Volcanic hydrothermal fluids might have transported mobile elements to the soil (Newsom and Hagerty, 1999). Large impacts played several key roles. They added volatiles but also removed them by impact erosion of atmosphere to space. Throughout much of Martian history, impacts might have episodically released volatiles to the surface and atmosphere (Segura et al., 2002) and created local hydrothermal systems (Newsom et al., 1996) that probably sustained habitable environments, much as volcanogenic hydrothermal systems have done throughout Earth history.

The Martian meteorites probably represent conditions of formation that are somewhat different from those that are presently found at the surface. Their composition is most consistent with alteration in an alkaline and reduced environment controlled by water-rock interactions rather than in the oxidized acidic environment at the surface (Zolotov and Shock, 2004). The carbonates in the ALH 84001 meteorite very likely derived from CO₂-rich fluids (Golden et al., 2001). However carbonates have not been documented to exist on the surface. Thermal infrared spectra obtained from orbit provide evidence that basalts in some areas have experienced
Figure 9. Schematic diagram of geochemical cycles of volatiles on Mars, showing reservoirs (boxes) in the mantle, crust, oceans and atmosphere, and showing the processes (arrows) that unite these reservoirs. Such cycles on Mars differ from those on Earth in several respects, including the following: Mars lacks standing bodies of water, atmospheric escape processes probably have exerted relatively larger effects, and subduction into the Martian mantle has been less important, if it occurs at all.

widespread alteration (Wyatt and McSween, 2003). Near-IR spectra from Mars Express indicate that clay minerals exist in the very ancient highlands (Bibring, 2005).

The Mars rover “Opportunity” has found clear evidence that aqueous processes at and near the Martian surface created chemical sediments (sulfates) and precipitates (hematite) in the Meridiani Planum region (Squyres and Knoll, 2005). The Spirit rover found evidence of aqueous basalt alteration and chemical precipitates in rocks on the floor of Gusev crater and in the nearby Columbia Hills (Ming et al., 2006).

Measurements of stable isotopes can reveal important aspects of the histories of volatiles in Martian materials. Each major volatile element has more than one stable isotope. Stable isotopes exhibit ranges of abundance ratios (e.g., D/H, $^{18}$O/$^{16}$O, $^{13}$C/$^{12}$C, $^{34}$S/$^{32}$S, $^{38}$Ar/$^{36}$Ar, etc.) in natural materials because isotopes having different masses can respond differently to chemical and physical processes. For ex-
ample, water molecules consisting of H and $^{16}$O have a higher vapor pressure than molecules having the heavier isotopes. When oxidized and reduced compounds of C (or S) coexist in chemical equilibrium, their lighter isotopes are relatively more abundant in the more reduced compounds. When organic matter is thermally decomposed, $^{12}$C-$^{12}$C bonds are broken more frequently, forming $^{12}$C-enriched gases.

The Martian atmosphere has been profoundly altered by the loss of volatiles to space due to a variety of processes (Kulikov et al., 2006). Hydrogen is lost by thermal escape. Photochemical reactions can deliver sufficient energy to heavier elements such as C, O and N to promote their escape. The loss to space of volatile compounds of these elements is expected to enrich the remaining atmosphere in the heavier isotopes D, $^{13}$C, $^{18}$O, and $^{15}$N (Jakosky and Jones, 1997). Gases in the upper atmosphere are not well mixed and so the lighter isotopes become relatively more abundant at higher altitudes where the probability of loss to space is greater. The Viking landers discovered that the Martian atmosphere exhibits relatively high values of D/H, $^{13}$C/$^{12}$C and $^{15}$N/$^{14}$N, and these high values are interpreted to reflect substantial losses to space of volatiles that were relatively enriched in H, $^{12}$C and $^{14}$N (e.g., Jakosky and Jones, 1997. These losses probably affected the long-term evolution of Martian climate and have contributed substantially to its present-day cold, dry state.

Martian meteorites have allowed geochemists to extend their measurements of isotopic patterns into the Martian crust and its potential reservoirs of volatiles. The isotopes $^{130}$Xe, $^{132}$Xe and $^{136}$Xe in gas trapped these samples display a range of relative abundances indicating that mixing has occurred between two end member reservoirs. The most likely explanation is that such mixing occurred on Mars; one of these end members is an atmospheric component enriched in the heavier isotopes. Meteorite mineral phases enriched in the lighter Xe isotopes are phases that probably formed earliest and that also retain Xe most tightly (Ott, 1988). Distributions of H isotopes in the Martian meteorites also indicate that multiple volatiles reservoirs exist (Watson et al., 1994). Elevated D/H values occur in minerals such as kaersutite and apatite that might have been influenced by exchange with D-enriched hydrogen in the atmosphere and in surface materials. The hydrogen reservoir having lower D/H values might consist of water and hydrated minerals from the crust and mantle that have not exchanged substantially with the atmosphere. Thus the lower D/H values might represent hydrogen from the interior of Mars that represents its primordial endowment of hydrogen, whereas elevated D/H values represent hydrogen reservoirs in surface materials and atmosphere that have been fractionated isotopically by processes of hydrogen escape to space that have been active for virtually all of Martian history. Values of $^{13}$C/$^{12}$C also provide evidence for multiple components of volatiles. Carbon components released at high temperatures from Martian meteorites during heating experiments have relatively low $^{13}$C/$^{12}$C values that are interpreted to represent the reservoir of C in the Martian mantle (Carr et al., 1985). In contrast, carbonates in these meteorites have
much higher $^{13}\text{C}/^{12}\text{C}$ values that have been interpreted to reflect exchange of C in the atmosphere and shallow deposits that has been fractionated isotopically by processes of atmospheric escape to space.

3.4. **Volatile and Life on Ancient Mars**

Cycling of these volatiles has evolved during Martian history. Although Mars and Earth are quite different today, their early evolution might have been similar because they shared similar processes that evolved in similar ways. For example, volcanism has been important on both planets and its intensity has declined over time. Both planets experienced early episodes of large impacts that were more substantial than later in history. And, of course, both planets share the same star and the effects of its long-term evolution. These processes had multiple consequences for the exchange of volatiles. Figure 10 depicts the processes that influenced the inventories of Martian volatiles and their exchange between the surface environment over time. A substantial endowment of volatiles on early Mars might have helped to maintain a denser atmosphere that, in turn, allowed liquid water to exist at the surface at least episodically and create the features described earlier by Sir Herschel. However over time these volatiles were removed either by substantial losses to space or by sequestration in the subsurface (Figure 3 in Jakosky and Jones (1997)).

![Figure 10. Schematic diagram of the history of key events and processes on Mars that have affected the cycling of volatiles between the interior, surface and atmosphere and their gains from and losses to space. Modified after Jakosky and Jones (1997).](image-url)
Under present harsh conditions at the Martian surface, the deep subsurface has probably provided the most widespread and stable environment for liquid water during Martian history (Clifford, 1993). Diverse populations of microorganisms (chemoautotrophs) that obtain their energy from redox reactions can populate subsurface environments on Earth (e.g., Kelley et al., 2005). Chemoautotrophs have been documented across a broad range of environmental extremes of temperature, pH and salinity (Rotschild and Mancinelli, 2001). Accordingly, chemoautotrophs that utilize \( \text{H}, \text{S} \) and \( \text{CH}_4 \) might represent useful analogs for potential Martian microbiota (Boston et al., 1992; Jakosky and Shock, 1998). The subsurface holds the greatest potential for having hosted the development of complex microbial ecosystems.

What does Mars tell us about the potential for habitable environments on other planets with masses different than that of Earth? Mars already has indicated that important geologic processes such as volcanism decline more rapidly on smaller planets than on Earth, leading perhaps to the more rapid deterioration of their surface environments. But how do planets more massive than Earth evolve over time? One might expect that a larger planet could maintain for longer periods of time the levels of geologic activity that were characteristic of Earth’s earliest history. Hydrothermal processes would be more widespread and processes of atmospheric escape would be less important on a planet larger than Earth. Future space exploration indeed promises new insights about the full diversity of habitable planetary environments and their potential to sustain life.

Guillame: Thank you Charles. Our final speaker is Charles Darwin. Mr Darwin’s inauspicious beginnings as a scholar included brief stints studying medicine and theology in Edinburgh and Cambridge, respectively. At Cambridge, his interest in the natural world was encouraged by John Henslow and Adam Sedgwick. The former arranged for his fateful voyage on the *Beagle* in 1831, and the rest, as they say, is natural history. Charles?

4. Co-evolution of the Atmosphere and Life on Earth

Darwin: Thank you, Charles. I will discuss how living things modify the geochemistry of the surface of the Earth, and how that affects atmospheric composition. Why the atmosphere? Earth’s atmosphere influences surface temperatures, and hence the availability of liquid water, by the greenhouse effect and by scattering sunlight back to space. It protects Earth’s biota from harmful radiation, it serves as a medium through which organisms exchange gases such as oxygen and carbon dioxide, and it is the principal reservoir of nitrogen at the surface of our planet. That the composition of the life-giving atmosphere is itself modified by biological activity gives rise to interesting feedbacks, as I shall discuss later. That modification is also an important signature for the presence of life, one that is potentially
detectable from distances as great as the nearest stars around which humans will search for, and may find, other Earth-like planets.

4.1. Evolution of the Atmosphere and the Evolution of Life

Science is certain that the composition of our atmosphere has dramatically evolved over our planet’s 4.5 billion year history, but past compositions are difficult to determine with any precision. The composition of the present atmosphere was only finalized after the discovery of argon by William Ramsay in 1894. The composition of Earth’s earliest atmosphere is the subject of much debate, as there is no rock record from this time, and the atmosphere may have been involved in the prebiotic chemistry that led to life (Miller and Urey, 1959). Perhaps the most significant change occurred within the first one hundred million years when any primitive atmosphere captured from the primordial solar nebula, or degassed during the impact of accreting planetary material was lost to space. Evidence for such atmospheric escape consists of the isotopic and elemental abundance of rare gases in the modern atmosphere (Porcelli and Pepin, 2000). These abundances are consistent with the pattern produced when heavier elements such as the rare gases are carried away by rapidly escaping hydrogen from the uppermost layers of the atmosphere. The hydrogen itself is heated to temperatures energetically equivalent to gravitational escape by elevated ultraviolet flux from the more active young Sun (Zahnle and Walker, 1982; Ribas et al., 2005) or by impacts.

The primordial atmosphere was replaced by one whose composition was determined by the chemistry of gases from volcanoes and fumaroles. If these gases exsolved from silicate melts that were in chemical equilibrium with the mantle, and if mantle chemistry has evolved little (Canil, 2002; Li and Lee, 2004), then these gases were primarily H₂O, CO₂, and N₂, with small amounts of CH₄ and H₂. Atmospheric composition continued to evolve, however, the most notable event being the appearance of persistent atmospheric oxygen about 2.4 billion years ago (Kasting, 1993). Evidence for anoxic conditions prior to that time now seems, I hesitate to say, airtight: That evidence consists of leaching of soluble ferrous iron (Fe²⁺) from paleosols (ancient soils), the presence of (insoluble) detrital pyrite (FeS₂) and uraninite (UO₂) in stream deposits, and mass-independent fractionation of stable isotopes of sulfur (Canfield, 2005). The last indicates the absence of the modern sulfur cycle (and, indirectly, the absence of atmospheric oxygen) in which sulfur dioxide from volcanoes or pyrite from sedimentary rocks is oxidized to sulfate in the oceans, and then re-reduced by sulfate-reducing microorganisms in marine sediments or the water column (Farquhar et al., 2000). Once oxygen appeared, we can say relatively little about its concentration, except that for the next billion years it was probably lower than today because the deep oceans remained anoxic (Kasting, 1987). Sometime in the late Precambrian, perhaps 700 million years ago, the level of oxygen rose to near its present value (Canfield and Teske,
1996) and has remained within a factor of two of that level since that time (Berner et al., 2000; Berner, 2001).

There is substantial, although not uncontroversial, evidence that atmospheric carbon dioxide concentrations were higher in the past (Rye et al., 1995; Pearson and Palmer, 2000; Kaufman and Xiao, 2003; Hessler, 2004). Decreasing carbon dioxide levels over geologic time might be symptomatic of a negative feedback that stabilizes climate against the increasing luminosity of the Sun on the main sequence (Walker et al., 1981). It is not known if the concentration of the primary component of the atmosphere, dinitrogen, has changed. The atmosphere is the principle reservoir of nitrogen on the surface of the Earth and modern nitrogen transformations are mediated by life, particularly bacteria. Fixation of nitrogen from the atmosphere into organic form, and its return to the atmosphere by biotic and abiotic processes cycles the entire atmospheric reservoir through living matter and sedimentary nitrogen in about 20 million years. In principle, the partial pressure of nitrogen could have evolved if a significant fraction of nitrogen had been shunted to these other reservoirs, but there is good reason to think this never occurred: The amount of nitrogen in living matter is limited by the total size of the biosphere and this has never been much larger in the past, as indicated by the relatively constant isotopic composition of carbon in marine carbonates (Schidlowsksi, 1988). The size of the sedimentary nitrogen budget has probably also has been limited by the propensity for life to efficiently scavenge sources of single nitrogen.

To discuss the effect of life on the chemistry of the atmosphere, one must discuss life, and to discuss life one must not only consider biochemistry, but history. Life has evolved not only according to absolute physiochemical limitations, but also, it seems, according to events that impose conditions on its subsequent evolution. Such contingencies are sometimes called ‘frozen accidents’. The small and large subunits of the ribosome, an ancient holoenzyme responsible for protein synthesis and critical to all known life, is one example of a possible frozen accident. The ribonucleic acid (RNA) polymer moiety of the ribosome, a possible relic from an early ‘RNA world’ that preceded the advent of proteins (Gilbert, 1986), has not been replaced by more efficient proteins as is presumably the case for other enzymes.

Speciation and the divergence of lineages of organisms is another example of a frozen accident. Once two species diverge sufficiently under the influence of natural selection they become genetically isolated and the process becomes irreversible, although it is now well established that even genetically distant microorganisms can exchange functional genetic material over evolutionary time through a process called horizontal gene transfer (Gogarten and Townsend, 2005). Despite this complication, the common ancestry of lineages, and the degree to which they have diverged, can be described in simplified terms as an evolutionary tree - a Tree of Life. This slide (Figure 11a) is a conceptual evolutionary tree that I drew in my journal. The next slide (Figure 11b) is a schematic of a phylogenetic tree constructed from the analysis of the sequences of conserved molecules in extant life.
Figure 11b can be used to illustrate several other ‘frozen accidents’ in the history of life: Most obvious is the existence of three “domains”, Bacteria, Archaea, and Eukarya (Woese and Fox, 1981), not two, or five, as previously thought (Whittaker, 1969). Another “accident” is the presence of the photosynthetic apparatus uniquely in the Bacteria. (Eukaryotic plants carry out photosynthesis because of the presence of chloroplasts - the extremely derived descendants of bacterial endosymbionts). Speaking of plants, all known algae are thought to originate from one or more endosymbiotic events in which an engulfed or infectious bacteria came to reside obligately in a eukaryotic host cell (McFadden, 2001). The eukaryotic cell may itself be the result of a fusion between two cell types, one of which gave rise to the nucleus (Gupta and Golding, 1996). The formation of the nucleus had consequences for the environmental range of the eukaryotic cell: The need to selectively transport messenger RNA out of the nucleus and proteins into the nucleus drove the evolution of large nuclear pore complexes. These large assemblies of proteins compromise the thermal stability of the nuclear membrane. As a result of this and other factors, eukaryotic cells have a much lower maximum known growth temperature, e.g., 62 degrees C for thermophilic fungi (Maheshwari et al., 2000) compared to as high as 121 degrees C in the nucleus-lacking prokaryotes (Kashefi and Lovley, 2003). I will discuss a conjectural consequence of this limitation for the atmosphere and global climate at the end of my talk.

Figure 11. (a) A reproduction of a phylogenetic tree from Charles Darwin’s “First Notebook on Transmutation of Species” (1837). (b) A schematic tree of life showing the phylogenetic relationships between some of the major kingdoms of organisms based on sequences of conserved molecules. The Bacteria and Archaea each comprise multiple, diverse groups that are not shown here for clarity. The location of the “root” of the tree in the Bacteria, inferred from ancient genes that duplicated before the appearance of the three domains, means that the Bacteria emerged as a distinct lineage before either the Archaea or Eukarya.
4.2. The Impact of Life on the Atmosphere

Life affects the atmosphere in three ways: First, organisms carry out chemical reactions directly involving atmospheric gases. Secondly, they mediate chemical reactions that indirectly involve atmospheric gases. Third, they create structures that alter the diffusion of gases between the atmosphere and surface of the Earth. Organisms interact directly with the atmosphere by producing and consuming gases, including N\textsubscript{2}, O\textsubscript{2}, CO\textsubscript{2}, CH\textsubscript{4}, and N\textsubscript{2}O. The last three molecules are strong greenhouse gases. Carbon dioxide is fixed by oxygenic phototrophs into complex organic molecules such as glucose (C\textsubscript{6}H\textsubscript{12}O\textsubscript{6}); oxygen is evolved as a by-product. The reverse reaction occurs when sugars are respired aerobically for energy (Table II). Carbon dioxide is also fixed or produced during chemosynthetic growth and catabolism of energy-rich organic molecules such as glucose under anaerobic conditions (fermentation), some reactions of which yield CO\textsubscript{2} and/or H\textsubscript{2}. Methane is produced either by combining CO\textsubscript{2} and H\textsubscript{2} or decomposing acetate (C\textsubscript{2}H\textsubscript{4}O\textsubscript{2}) (Table II). Methane is consumed during aerobic or anaerobic methane oxidation. The latter takes place concomitantly with the reduction of sulfate (SO\textsubscript{4}\textsuperscript{2-}) to sulfide (HS\textsuperscript{-}) in marine environments. (Sulfide evolves to the gas H\textsubscript{2}S under low pH). Nitrous oxide is produced during denitrification by the reduction of nitrate (NO\textsubscript{3}\textsuperscript{-}) to nitrite (NO\textsubscript{2}\textsuperscript{-}) and nitric oxide (NO). Dinitrogen (N\textsubscript{2}), the dominant constituent of Earth’s modern atmosphere, is converted into organic form by the process of nitrogen fixation and also released during denitrification.

<table>
<thead>
<tr>
<th>Metabolism</th>
<th>Reaction</th>
<th>Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photosynthesis</td>
<td>6CO\textsubscript{2} + 6H\textsubscript{2}O → C\textsubscript{6}H\textsubscript{12}O\textsubscript{6} + 6O\textsubscript{2}</td>
<td>B, E</td>
</tr>
<tr>
<td>Aerobic respiration</td>
<td>C\textsubscript{6}H\textsubscript{12}O\textsubscript{6} + 6O\textsubscript{2} → 6CO\textsubscript{2} + 6H\textsubscript{2}O</td>
<td>B, A, E</td>
</tr>
<tr>
<td>Sugar fermentation</td>
<td>C\textsubscript{6}H\textsubscript{12}O\textsubscript{6} → 2C\textsubscript{2}H\textsubscript{5}OH + 2CO\textsubscript{2}</td>
<td>B, E</td>
</tr>
<tr>
<td>Methanogenesis</td>
<td>CO\textsubscript{2} + 4H\textsubscript{2} → CH\textsubscript{4} + 2H\textsubscript{2}O</td>
<td>A</td>
</tr>
<tr>
<td>Acetoclastic methanogenesis</td>
<td>CH\textsubscript{3}COOH → CO\textsubscript{2} + CH\textsubscript{4}</td>
<td>A</td>
</tr>
<tr>
<td>Methane oxidation</td>
<td>CH\textsubscript{4} + 2O\textsubscript{2} → CO\textsubscript{2} + 2H\textsubscript{2}O</td>
<td>B</td>
</tr>
<tr>
<td>Anaerobic methane oxidation</td>
<td>CH\textsubscript{4} + SO\textsubscript{4}\textsuperscript{2-} + 2H\textsuperscript{+} → H\textsubscript{2}S + CO\textsubscript{2} + 2H\textsubscript{2}O</td>
<td>A</td>
</tr>
<tr>
<td>Nitrogen fixation</td>
<td>N\textsubscript{2} → N\textsubscript{org}</td>
<td>B,A?</td>
</tr>
<tr>
<td>Denitrification</td>
<td>2NO\textsubscript{3} \textsuperscript{-} + 2H\textsuperscript{+} + 5H\textsubscript{2} → N\textsubscript{2} + 6H\textsubscript{2}O</td>
<td>B, A</td>
</tr>
<tr>
<td>Ammonium oxidation</td>
<td>2NH\textsubscript{4} \textsuperscript{+} + 4O\textsubscript{2} → 2NO\textsubscript{3} \textsuperscript{-} + 4H\textsuperscript{+} + 2H\textsubscript{2}O</td>
<td>B</td>
</tr>
<tr>
<td>Anammox</td>
<td>NH\textsubscript{4} \textsuperscript{+} + NO\textsubscript{5} \textsuperscript{-} → N\textsubscript{2} + 2H\textsubscript{2}O</td>
<td>B</td>
</tr>
<tr>
<td>Sulfate reduction</td>
<td>SO\textsubscript{4} \textsuperscript{2-} + 2H\textsuperscript{+} + 2CH\textsubscript{2}O → H\textsubscript{2}S + 2CO\textsubscript{2} + 2H\textsubscript{2}O</td>
<td>B,A</td>
</tr>
</tbody>
</table>

The distribution of these metabolisms in the tree of life is very uneven (Table II). Although organisms capable of aerobic respiration are present in all three domains...
of life (Bacteria, Archaea, and Eukarya), only a single group of the Bacteria - the Cyanobacteria - are known to be capable of oxygenic photosynthesis. Likewise, methanogens are found only amongst the Archaea. The genes for proteins involved in nitrogen fixation have been found in both the Bacteria and Archaea (particulary methanogens) (Zehr et al., 2003). By comparison, eukaryotic cells are not very metabolically facile, and so far shown only to be capable of aerobic respiration, oxygenic photosynthesis, and fermentation. In fact, the first two metabolisms were imported long ago when oxygen-respiring α-proteobacteria and cyanobacteria came to permanently occupy the eukaryotic cell and the eukaryotic plant cell, eventually becoming the mitochondrion and the chloroplast, respectively. Methanogenic symbionts have been found in certain protists, but it is not known if these benefit the eukaryotic host cell (Fenchel and Finlay, 1991).

Life also alters atmospheric composition indirectly by mediating chemical reactions that affect atmospheric gases. One example is biological control over weathering reactions of silicate minerals at the surface of the Earth. Many silicate minerals created at high pressures and temperatures within the Earth are unstable under conditions at the Earth’s surface. In the presence of water and an acid they transform to more thermodynamically favored phyllosilicates, clays and oxides by a variety of weathering reactions, e.g.,

$$\text{CaAl}_2\text{Si}_2\text{O}_8 + 8\text{H}^+ \rightarrow \text{Ca}^{2+} + 2\text{Al}^{3+} + 2\text{SiO}_2 + 4\text{H}_2\text{O}. \quad (2)$$

Protons can be supplied by carbonic acid, $\text{HCO}_3^-$, a ubiquitous acid in waters under CO$_2$-containing atmospheres;

$$\text{CO}_2(aq) + \text{H}_2\text{O} \rightarrow \text{H}_2\text{CO}_3 \rightarrow \text{HCO}_3^- + \text{H}^+. \quad (3)$$

Abiotic weathering of silicates and its role in atmospheric CO$_2$ and climate have been reviewed by Kump et al. (2000). Plants and fungi produce a variety of organic acids such as oxalic acid that accelerate the weathering process. Vegetation may also increase the relative mobility and release of Ca, Mg, and Sr, thereby favoring the right-hand side of Equation 2 (Gislason, 1996). (Vascular plants also influence weathering in other ways, as I shall talk about in a moment). Acceleration of weathering by organisms may have counteracted warming by a brighter, evolving Sun (Schwartzman and Volk, 1989; Berner, 1992).

The complementary reaction to weathering is the biogeic precipitation of aragonite or calcite in the oceans;

$$\text{Ca}^{2+} + 2\text{HCO}_3^- \rightarrow \text{CaCO}_3 + \text{CO}_2 + \text{H}_2\text{O}. \quad (4)$$

In the inorganic cycling of carbon, net formation of sedimentary carbonates balances CO$_2$ produced by volcanoes and metamorphism. Photosynthetic uptake of carbon dioxide favors carbonate precipitation, however the rate will still be limited by the supply of calcium ions to the ocean. For billions of years before the emergence of multicellular life, precipitation was abiotic or driven by photosynthesis.
Biomineralization by animals involves actively maintaining and controlling the precipitation process. Active pumping of calcium ions and bicarbonate to biomineralization sites means that ambient concentrations, and thus the calcite saturation state of the ocean, can fall. Equation 3 tells us that ambient CO$_2$ concentrations will also fall until equilibrium is re-established. This has obvious consequences for climate, and the suggestion that animals were somehow responsible for a dramatic cooling of the planet at the end of the Precambrian is not new. However, the oldest fossils of biomineralizing animals, the “small shelly fauna” including *Cloudina* and *Namacalathus*, appear 550 million years ago (Grant, 1990; Grotzinger et al., 2000), and only after multiple episodes of extreme glaciation and the decline of massive limestone formations.

Biotic manipulation of atmospheric composition can also affect the fraction of sunlight that is reflected back into space, a quantity called the planetary albedo. Biogenic aerosols influence climate by scattering incoming radiation and serving as cloud condensation nuclei (CCN). Natural aerosols are primarily composed of sulfate and carbonaceous compounds (Pöschl, 2005) and, in the marine boundary layer, salts (Murphy et al., 1998). Biogenic sulfate derives from oxidation of dimethyl sulfide (DMS) from marine organisms and carbony sulfide (COS) from terrestrial vegetation (Andreae and Crutzen, 1997). An elevated atmospheric CCN content decreases average cloud droplet size, increasing cloud albedo and cloud lifetime against precipitation, thereby cooling the Earth (Ackerman et al., 2000). Aerosols may have been more abundant on the Archean Earth when oxygen was absent from the atmosphere (Pavlov et al., 2001b).

The final effect of living organisms is what one might call a “compartmentalization” of the atmosphere by biogenic structures. One example of this type of effect is a soil: Soils are an admixture of the weathering products of rocks and decomposing organic matter. Physically, a soil is a permeable structure, maintained in place by vascular plant roots and its own cohesiveness. It contains significant pore space containing gases and aqueous solutions and because their transport is limited by diffusion, their composition can depart radically from that of the atmosphere. These differences can affect the pathways and kinetics of the weathering reactions of the parent rock and the decomposition of organic matter. It will also affect biological activity through, for example, the abundance of oxygen. More specifically, aerobic respiration of organic matter by the roots and soil microorganisms can increase the concentration of CO$_2$ in soil pore space to as much as 10,000 ppm. There is also a concomitant decrease in the concentration of O$_2$. Because there is much more O$_2$ in the atmosphere compared to CO$_2$, the decrease in O$_2$ will in general not be significant. The exceptions are soils in which a significant amount of the pore volume is occupied by water, through which oxygen diffuses a factor of 10$^4$ times more slowly than in air, and in which conditions can be hypoxic or even anoxic. High soil CO$_2$ leads to an increase in the concentration of hydrogen ions in porewater (Equation 3), hence favoring chemical weathering (Equation 2). In the absence of soil, atmospheric CO$_2$ and, through the greenhouse effect, temperatures would be
much higher (Berner, 1992). Vascular plans also actively pump soil water into their tissues and transpire it into the atmosphere, thereby changing relative humidity, cloud formation, and planetary albedo in unknown but potentially significant ways.

A second example is the digestive tract of animals. Rather than simply filtering particles from seawater (as sponges do) or absorbing dissolved molecules (which the enigmatic placozoans do), animals with guts process their food in internal compartments that are isolated from the atmosphere. Again, as in the case of soils or sediments, biological activity by the microorganisms, in addition to the host animal itself, alters the chemistry of the immediate environment compared to surface conditions at the Earth or in the oceans (Plante and Jumars, 1992). To illustrate how dramatic the difference is, I show here a table (Table III) giving median concentrations of gases in human flatus (Suarez et al., 1997) compared with the composition of the modern atmosphere. Organic matter passing through guts is processed under conditions quite dissimilar to those at the surface of the modern Earth: Metabolisms such as methanogenesis that require abundant H$_2$, for example, are favored. The consequences for the global carbon cycle of the appearance with animals with guts are not understood but may have been dramatic (Rothman et al., 2003).

Compartmentalization of the atmosphere did not stop with the origin of animal guts: New forms emerged as animals evolved, including burrows, reefs, and termite mounds. We humans continue a time-honored tradition when we build and inhabit edifices, heating or cooling their interiors by burning fossil fuels, releasing CO$_2$, and altering the outside atmosphere and climate.

<table>
<thead>
<tr>
<th>Chemical species</th>
<th>Flatus</th>
<th>Present Atmosphere</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO$_2$</td>
<td>29.5%</td>
<td>0.038</td>
</tr>
<tr>
<td>H$_2$</td>
<td>32.2</td>
<td>0.00005</td>
</tr>
<tr>
<td>N$_2$</td>
<td>18.9</td>
<td>78</td>
</tr>
<tr>
<td>CH$_4$</td>
<td>0.006</td>
<td>0.00019</td>
</tr>
<tr>
<td>O$_2$</td>
<td>2.95</td>
<td>21</td>
</tr>
<tr>
<td>H$_2$S</td>
<td>1.1</td>
<td>0</td>
</tr>
</tbody>
</table>

4.3. FEEDBACKS AND PLANETARY HOMEOSTASIS

I have described three ways in which living things alter the atmospheric composition of the Earth, and thereby the climate and environment which they experience. I have also described in previous work (Darwin, 1859) how the environment acts
through natural selection on the variation between individuals in a species, thereby causing the species to evolve, or even new species to emerge, over geologic time. Does this combination of forces constitute a feedback between the organic and inorganic worlds? Margulis and Lovelock (1974) proposed that surface conditions are regulated by life so as to permit maximum growth of the biosphere and that planetary homeostasis emerges naturally. This extraordinary claim (the “Gaia hypothesis”) has met with great interest, and perhaps even greater incredulity by the scientific community.

Watson and Lovelock (1983) illustrated planetary homeostasis by a simple model involving oppositely-pigmented members of the Asterales order of the Plantae (black and white daisies). Both strains have the same growth response to temperature, but because they absorb sunlight to a different degree (black completely, white not at all), they experience different local temperatures. As a result, the growth of one type of daisy will be favored until that growth has changed the planet’s albedo and has heated or cooled it sufficiently to eliminate that advantage. As a consequence, the evolving brightness of the planet’s parent star will be accompanied by a concomitant change in the relative proportions of black and white daisies. Black daisies will be favored at lower illumination levels; white daisies at higher.

The surface temperature of Daisyworld remains constant as an initially pure black daisy population is replaced by white daisies (Figure 12a). Subsequent versions of the Daisyworld model included intermediate strains of “grey” daisies and predation (Harding & Lovelock, 1996).

Evolutionary biologists have been reluctant to take such models seriously because they seem to imply phenomenae that would not emerge from natural selection acting on individual organisms. In particular, the idea that natural selection can act on groups of organisms (“group selection”), rather than individuals, has fallen out of favor, with the exception of close relatives that share alleles (“kin selection”). Another objection is that adaptation of species to their environment should lead to an uncontrolled “drift” of the system and loss of self-regulation (Robertson and Robinson, 1998; Lenton, 1998). Indeed, it would seem that the essence of Daisyworld is lack of or incomplete adaptation to the environment due to some physiological barrier, for example. In the case of the original model, it is the finite temperature range of growth (Lenton and Lovelock, 2000).

Does Earth’s atmosphere reflect any Daisyworld-like self-regulation? Charlson et al. (1987) have proposed that the production of dimethylsulfide by marine algae, the incorporation of that sulfur into cloud condensation nuclei and the effect of clouds on the global albedo is one such atmospherically-mediated climate regulation. There are many other potential climate-biosphere feedbacks (Lashof et al., 1997). One such feedback may involve the maximum growth temperatures of lichens and plants involved in enhancing weathering of silicate rocks. The upper temperature range for growth is about 62°C for fungi (Maheshwari et al., 2000) and 40-50°C for plants. (Lower limits are set by the unavailability of significant liquid water below 0°C.) As a result, the latitude range of plants shifts in response to
climate (Davis and Shaw, 2001). Thus, if the concentration of atmospheric carbon
dioxide decreases, temperatures fall, plants retreat from high latitudes and weather-
ing at those latitudes decreases. As a consequence, carbon dioxide levels increase,
warming the Earth. Conversely, if the Earth warms, plants occupy higher latitudes,
increasing the development of soil and rates of weathering there, and lowering
atmospheric CO$_2$. This feedback is in addition to the abiotic thermostat proposed
by Walker et al. (1981). The vegetation migration effect is stronger than the abiotic
effect alone at low atmospheric pCO$_2$ levels and temperatures, but becomes weaker
with increasing CO$_2$ and mean surface temperature because high latitudes consti-
tute a smaller fraction of Earth’s weatherable surface area (Figure 12b). A potential
instability point exists when temperatures at the equator approach the limit of plant
growth. Beyond that point, plants die back at low latitudes, leading to decreased
weathering, increased CO$_2$, and still higher temperatures. A similar collapse of
regulation occurs in a 2-D Daisyworld model when increasing equatorial temper-
atures drive desertification and an increase in albedo at low latitudes (Ackland et
al., 2003).

Figure 12. (a) Biological homeostasis ("Daisyworld") in the abstract: As an external parameter (e.g.,
stellar radiation) varies, the relative growth of two species (black and white daisies), which affect
the internal state (e.g., temperature) in opposite ways, causes the internal state to remain constant
over a certain range (solid curve). Curves of growth rate as a function of the internal parameter value
are at the top; the relative population sizes are the bottom curves. Regulation fails once one species
dominate. (b) One possible example of climate regulation by life: Relative rate of weathering and
removal of CO$_2$ from the atmosphere as carbonates as a function of atmospheric CO$_2$, based on
a coupled climate-vegetation-soil model (Gaidos, in prep). The solid curve is for the Earth with
the current configuration of continents, orbital parameters, and fixed vegetation biomass; the dotted
curve is for an unvegetated world. The steeper slope of the vegetated world is due to migration of
vegetation to, and soil development at, high latitudes with increasing temperature as well as enhanced
gross primary productivity under a CO$_2$-rich atmosphere. A steeper slope implies a stronger negative
feedback.
The Gaia hypothesis emerged from the proposition that life on Mars should have already been detected by its effect on atmospheric composition, and therefore did not exist in abundance (Hitchcock and Lovelock, 1966). Regardless of what one thinks of the Gaia conjecture, the notion that Earth’s atmosphere bears the imprint of life continues to be accepted as a practical approach to detecting life on another planet and recent studies have addressed this possibility (Des Marais et al., 2002). The simultaneous presence of oxygen and methane in an atmosphere, originally considered by Hitchcock and Lovelock (1966), is considered to be a strong indicator of biological activity and in an interesting experiment, they were detected in Earth’s atmosphere by a passing spacecraft (Sagan et al., 1993).

Since my publication of *Origin of Species*, one species in particular has increased its influence on the atmosphere. The atmospheric concentration of CO$_2$ is now at the highest level it has ever been in the past 600,000 years (Siegenthaler et al., 2005) and the annual increase shows no signs of abating (Keeling et al., 1989). The concentration of methane has also dramatically increased since the expansion of major anthropogenic sources, particularly rice paddies and ruminent livestock. Our influence on the atmosphere would be detectable from far away, if an observer knew what to look for. Our engineering of the atmosphere is reckless and without a plan. Intelligent species on other planets might be more deliberate and constructive in their work. For example, they could introduce a suite of long-lived “super” greenhouse gases such as fluorine compounds that would block most of the infrared radiation escaping from the surface, creating an intense greenhouse effect that would warm planets like Mars that are too distant from their parent star and otherwise too cold to support liquid water at their surfaces (Gerstell et al., 2001). The unambiguous spectroscopic signature of such a feat - strong absorption features at many infrared wavelengths - might be detected from a great distance. Thank you.

5. Discussion

Guillaume: Thank you, Charles. We now move into the panel discussion part of our session. Charles, I see you already have your hand up, so please go ahead.

Lyell: Are there examples of destabilizing, or “anti-Gaia” biotic feedbacks? The geologic activity of our planet is also thought to have slowly decreased with time as the radiogenic heat produced in the Earth decays and the mantle convection needed to reject that heat slows. What are the implications of this for climate stabilization by biology?

Darwin: Excellent question. Lovelock and Kump (1994) addressed some aspects of this issue, arguing that above a global mean temperature of 20°C, both marine and terrestrial ecosystems supply a destabilizing positive climate feedback by taking up less CO$_2$ or releasing more to the atmosphere. Clearly, long-term bright-
ning of the Sun will eventually defeat any stabilization. Kasting and Caldeira (1992) showed that, well before the increase in surface temperature and atmospheric water vapor content create a runaway “wet” greenhouse, the atmospheric carbon dioxide concentration will fall below the level where C4 plants can carry out photosynthesis. Of course, strong selection pressure may produce plants capable of growing at even lower CO₂ concentrations but eventually levels will reach zero and a biotic crisis will ensue. Lower rates of geologic activity and CO₂ output in the future (Franck et al., 2000) will only hasten that day.

**Darwin:** I have a question for Sir Herschel: You mention that greenhouse gases other than carbon dioxide may have been important in determining the climate and of early Mars. What are those gases?

**Herschel:** The most likely candidate is methane. Per molecule, methane is twenty times more effective as a greenhouse gas. Methane is an attractive candidate because it absorbs at different wavelengths than CO₂. Kasting (1997) and Pavlov et al. (2001a) argued that an atmosphere with about one bar of CO₂ and 1% methane could keep Mars surface temperatures above freezing early in its history when the Sun was fainter. The real difficulty with this scenario is that it requires a source of CH₄ that is comparable with the present-day biological source on Earth (much of which is anthropogenic in nature). Purely abiotic geologic activity produces methane on the Earth, i.e., through production of H₂ by the serpentinization of mafic rocks and reaction of hydrogen with CO₂, but in relatively small (and uncertain) amounts compared to biogenic methane. Ongoing fluid-rock chemistry in the crust may be responsible for the purported few tens of ppb of CH₄ detected on Mars (Lyons et al., 2005). Alternatively, methane may only now be escaping from the destabilization of hydrates formed during an earlier, more geologically active phase (Oze and Sharma, 2005). However, the level of geologic activity on early Mars is not known and it is not at all clear that even a very active Mars could have produced the necessary methane to warm the surface without the intervention of biology. Furthermore, the photochemical lifetime on an ancient, wet Mars would be much shorter than its present value of a few centuries. So perhaps this all implies that life - methanogens - did exist on Mars!

**Darwin:** My question is for Sir Lyell: After the “Opportunity” and “Spirit” rovers, which question about the past or present habitability of Mars now looms largest? What are the next “opportunities”, if I may turn a phrase, to answer that question?

**Lyell:** The twin Mars Exploration rovers have demonstrated that liquid water once existed at and near the surface and that it chemically altered the rocks. But spacecraft orbiting Mars have revealed other extensively altered terrains where environments might have been even more hospitable to life and persisted for longer periods. The presence of methane in the Martian atmosphere might indicate that
liquid water exists even today in the subsurface. Every mission reveals that the Martian crust and atmosphere are even more complex than we had previously suspected. Therefore the next question looms largest: Did liquid water ever co-exist with sources of energy and environmental conditions that could have allowed life to develop and persist? Additional questions immediately follow. What has been the full diversity of habitable environments on Mars and can spacecraft explore them and/or their geologic records? How extensive geographically were such environments and how long did they persist? We must identify geologic deposits that preserved evidence of ancient habitable environments and also, potentially, signatures of life. Examples of such deposits include carbonates, aqueously deposited silica, phosphates, and evaporite minerals such as well-cemented sulfates and halides.

Regarding upcoming opportunities, both the Mars Express and the Mars Reconnaissance orbiters will probably continue to identify minerals that indicate where volatile species such as water and CO$_2$ have interacted with crustal rocks and perhaps created both habitable environments and aqueous precipitates. The orbiters will also probably discover sites that are even more promising than those visited by the MER landers. Both NASA’s Phoenix lander (2007 launch) and Mars Science Laboratory (MSL) rover (2009 launch) will analyze both minerals and volatile species. The MSL rover, another ”robotic field geologist”, like “Spirit” and “Opportunity”, might someday explore a site that represents an ancient lake, hydrothermal system, or some other kind of ancient habitable environment that, once upon a time, indeed fulfilled all of life’s requirements.

Guillaume: This is indeed a fascinating discussion, but my highly accurate Swiss watch tells me that we are out of time. Let’s thank our speakers again and enjoy the rest of this day in this especially habitable part of Earth.

Acknowledgements

DJD acknowledges support from the NASA Astrobiology Institute and the NASA Exobiology program. EG acknowledges travel support by the NASA Terrestrial Planet Finder Foundation Science program and the NASA Astrobiology Institute. This chapter was improved by reviews by S. Rodin, L. Ksanfomality, and K. Fishbaugh.

References

Christensen, P. R.: 2003, Nature 422, 45.

*Address for Offprints:* E. Gaidos, University of Hawaii, 1680 East-West Road, Honolulu, Hawaii 96822 USA. gaidos@hawaii.edu