The rise of oxygen and the hydrogen hourglass

Kevin J. Zahnle, David C. Catling, Mark W. Claire

1. Introduction

An oxygen atmosphere raises two issues that are not always separated. One is the matter of abundant \( \text{O}_2 \), which is the distinctive feature of Earth’s atmosphere. The other is the oxidation of the surface and atmosphere, a state that is widespread in the solar system. Oxygen and oxidation are different things and reflect different processes acting on different time scales, although it is plausible that one is prerequisite to the other. It could be that abundant free oxygen in the atmosphere oxidized the surface, or it could be that oxidation of the surface allowed abundant free oxygen to endure. The view taken here is that surface oxidation is prerequisite to \( \text{O}_2 \). It has long been considered probable from hints in the geological record that oxygenic photosynthesis appeared much earlier than widespread crustal oxidation (Holland, 1962; Buick, 2008), and thus that surface oxidation played a role in the rise of oxygen (e.g., Berken and Marshall, 1965). Where we differ is that we think of oxygen as a consequence of the steady loss of hydrogen to space. Moreover, we explicitly tie the time it took to oxidize the surface through hydrogen escape to the time taken between the origin of Earth’s atmosphere and the advent of oxygenic photosynthesis, the two necessary conditions for oxygen in the atmosphere.
of oxygenic photosynthesis and the establishment of an oxygen atmosphere (Catling et al., 2001; Claire et al., 2006).

2. Oxidation

Sixty years ago Urey (1952) wrote that “the highly oxidized condition is rare in the cosmos and exists in the surface regions of the earth and probably only in the surface regions of Venus and Mars. Beyond these we know of no highly oxidized regions at all, although undoubtedly other localized regions of this kind exist.” His underlying interest in the matter is revealed in his next sentence, “This is essentially the argument of Oparin.” To Oparin and Urey the reduced atmosphere is to be expected at early times relevant to the origin of life, while the highly oxidized condition of planetary surfaces is an evolutionary result of hydrogen escape. Like Oparin, Urey regarded a reduced atmosphere as essential to prebiotic synthesis, but he also regarded the tension between the reduced interior and the oxidized surface as contributing to life's subsequent evolution. We agree.

Space age exploration has shown that the highly oxidized condition is somewhat more common than Urey expected it to be. Oxidation of planetary surfaces is not rare in the solar system. In all cases apart from Earth the oxidation is clearly caused by hydrogen escape. For example, several icy satellites (Ganymede, Europa, Rhea) have extremely thin O2 atmospheres derived from splitting water molecules in ice by uv photons or energetic particle bombardment (Cruikshank, 2010). Hydrogen escapes easily but oxygen, which is much heavier, does not. As a consequence over time the surface ice and any contaminants in the ice become highly oxidized.

Mars provides a more Earth-like example closer to home. Mars is red because much of the iron at its surface is oxidized. On Earth surfaces like these (“redbeds”) first appear in the geological record during the Huronian glaciation ca. 2.32 Ga and are one of the classic indicators of an oxic atmosphere (Holland, 1999; Kump, 2008; Guo et al., 2009). Mars's surface is locally characterized by the presence of strong oxidants such as peroxides (seen by Viking, Huntken, 1979) and perchlorates (seen by Phoenix, Hecht et al., 2009), and on a global scale there are extensive and locally thick sulfate deposits (encountered by the Mars Exploration Rovers on the ground and mapped by satellites from above). Hydrogen escape is fast enough on Mars to generate its modest atmospheric reservoir of O2 in just 1013 years (Nair et al., 1994; Zahnle et al., 2008). On Mars the oxidation appears to be quite shallow, with no evidence that oxidation extends to the mantle or even to the deeper crust. If so, planetary oxidation would have been quick (Hartman and McKay, 1995).

In contrast to Mars, the most modern accounts of the rise of oxygen on Earth either marginalize or ignore the role of hydrogen escape in oxidizing the surface. Rather, oxidation of the surface is usually ascribed to a shrinking influence of reduced volcanic gases from a reduced mantle, measured relative to burial of reduced carbon (from CO2) in continents, although the detailed mechanism is debated (Kasting et al., 1993; Kump et al., 2001; Holland, 2002, 2009; Kump and Barley, 2007; Gaillard et al., 2011). The driving force is planetary cooling. Hydrogen escape plays a subsidiary role in these models: it is determined by the difference between mantle degassing and carbon burial, subject to the constraint that the Archean not become too reduced. A variant that is closer to our point of view posits preferential subduction of reduced matter by the mantle (Hayes and Eldridge, 2006). The variant is easier to understand because, like hydrogen escape, if it worked it would actually oxidize the surface.

It was long hoped that the answer would be provided by a gradual oxidation of the mantle as it aged (Kasting et al., 1993; Holland, 2002), but evidence that the mantle's oxidation state has not changed in any measurable way since 3.8 Ga (Delano, 2001; Canil, 2002; Li and Lee, 2004), a state that may even extend as far back as 4.35 Ga (Trail et al., 2011), has been received as convincing. More recent versions of the hypothesis compare the steady reducing power of volcanic gases to a secular increase in the amount of carbon and sulfur that these gases would have to reduce (Holland, 2009), or to a secular decrease in the reducing power of volcanic gases by eroding them at later times under shallow water or directly into the atmosphere from emergent continents rather than at high pressure under the deep ocean (Kump and Barley, 2007; Gaillard et al., 2011).

But, stepping back from the details, the idea that the mantle should oxidize the surface is rather puzzling. The gases that come from the mantle are reducing and have always been reducing, because the mantle is reduced. That they might have been more or less reducing in the past does not make them oxidizing. The reasonable expectation is that the surface would be reduced by these gases, or by any other interaction with the mantle. Net oxidation of the surface by the mantle can occur only if the surface exports more gruntant to the mantle (via subduction) than the mantle exports via volcanic gases. In other words, convention makes the rather awkward argument that an oxidized surface oxidizes itself by further reducing an already reduced mantle. The Hayes and Eldridge mechanism is not immune from this criticism. By contrast hydrogen escape is categorically oxidizing.

3. Oxygenic photosynthesis

There is at present little doubt that oxygenic photosynthesis is required to create an oxygen-rich atmosphere like Earth's. Although it is possible to imagine atmospheres that might become oxygenated abiotically through some combination of vigorous hydrogen escape and high rates of stellar ultraviolet irradiation, to the best of our knowledge no one has yet succeeded in demonstrating this with an actual photochemical model.2 The thermodynamics of typical planetary materials favor O2 only at very high temperatures. Photosynthesis makes O2 from 5800 K sunlight. Even under Mars's favorable circumstances — weak gravity, an oxidized surface, extremely limited weathering owing to extremely low water activities, low surface temperatures, and minimal volcanism and crustal recycling — abiotic processes produce an O2 partial pressure that is only 4 × 10−5 that of Earth. Detailed photochemical models show that for an Earth-like planet the abiotic photochemical source of O2 would be overwhelmed by Earth's volcanic gases, resulting in O2 partial pressures at the surface on the order of 10−12 bar (Kasting and Walker, 1981; Kasting, 1993; Haqq-Misra et al., 2009).

Oxygenic photosynthesis uses sunlight to split water molecules into hydrogen and oxygen. The O2 is released to the atmosphere as a by-product. The hydrogen is used to reduce CO2 to water and organic matter. Under an oxygenated atmosphere most of the organic matter is aerobically respiried, closing the cycle, but a small fraction of the organic matter is further reduced to inedible carbon that is ultimately buried. Methane is also a product but it is mostly eaten (using O2 or sulfate) before it reaches the atmosphere. On long time scales the net effect of oxygenic photosynthesis today is approximately CO2 → C + O2. This is why carbon burial is often treated as the source of O2 (Hayes et al., 1983). Many models focus on biological and geological controls on carbon burial to reconstruct Phanerzoic histories of O2 (Berner, 2001).

Under an anoxic atmosphere, aerobic respiriation might still be possible locally where O2 is made, but we would expect a bigger role for anaerobic pathways and a higher fraction of the organic matter to be reduced, and a much higher fraction of the organic matter made by oxygenic photosynthesis to reach the atmosphere as methane (Walker, 1987; Catling et al., 2001; Kasting and Sieffert, 2002).

2 It has very recently been shown in unpublished work (by us and by others) that it is possible to generate an O2-rich atmosphere abiotically by photochemistry of an Earth-like planet in the habitable zone of a very uv-quiet M dwarf. A distinctive and inevitable feature of these hypothetical atmospheres is that they are also CO-rich: they have roughly twice as much CO as O2, because both gases come from CO. It is our opinion that such an atmosphere does not resemble Earth's. Technically-speaking, it is not even oxidized. It is also flammable.
Most of the CH₄ reaches the stratosphere, where it is oxidized with the help of solar UV. Most of the hydrogen that had been in the CH₄ is destined to escape to space. The net is therefore H₂O → O₂ + H₂ (→ space). When oxygenic photosynthesis leads to methane, hydrogen escape becomes a source of O₂ (Catling et al., 2001).

Methane is much less reactive than O₂. If both are vented to the atmosphere by an oxygenic photosynthetic ecosystem, the O₂ and free radicals that are made from O₂ will react with rocks and such (Berkner and Marshall, 1965) while the CH₄ reacts only with free radicals generated from O₂. As a result one would expect the atmosphere after the origin of oxygenic photosynthesis to be CH₄-rich rather than O₂-rich, and that this would remain the state of the atmosphere for as long as chemical sinks for O₂ remained easy and plentiful. Methane itself leaves few geological traces, but it can polymerize to soots, possibly nitrogenous or sulfurous (Pavlov et al., 2001; Trainer et al., 2006; Domagal-Goldman et al., 2008; Zahnle, 2008; Zerkle et al., 2012), that should be isotopically and structurally distinctive. A methane-rich atmosphere also welcomes other more fragile biogenic reduced gases, like sulfides, and thus may reveal itself indirectly through its influence on the sulfur cycle, a topic we have speculated on elsewhere (Zahnle et al., 2006).

It is sometimes argued that the advent of oxygenic photosynthesis was a sufficient condition to create an O₂ atmosphere (e.g., Kopp et al., 2005). The appeal here is to brute force: photosynthesis using H₂O is presumed to be so much more productive than photosynthesis reliant on other hydrogen donors that it ought to take over the world on a biological time scale. This is not an irrefutable notion (cf. Sleep and Bird, 2008). It is not obvious a priori that the hydrogen source was enormously more limiting to biological productivity than all other nutrients (cf. Kharecha et al., 2005). A second presumption is that the flux of biogenic O₂ into the atmosphere would be great enough to utterly overwhelm the capacity of weathering reactions or hungry mouths to consume it; this too is at best debatable. But the bigger problem with this hypothesis is not its logic but rather that it has not proved possible to identify exactly when in the geologic record the one true revolution took place (Holland, 1962; Des Marais, 2000; Buick, 2008; Guo et al., 2009). One extreme solution to this paradox is to argue that Earth’s atmosphere has always been oxygenated and that the origin of oxygenic photosynthesis predates the geologic record (Ohmoto, 1999). However, the overwhelming consensus recognizes several substantial changes in redox conditions at Earth’s surface between 3.5 Ga and 2.1 Ga, to say nothing of the evolution of multicellular animals ~0.6 Ga (Yin et al., 2007; Love et al., 2009; Malloof et al., 2010), an innovation often explained as heralding the advent of breathable levels of O₂ (Knoll, 2004).

4. Where does oxygen come from?

Excess oxygen in the crust and atmosphere derive ultimately from splitting H₂O followed by H escape to space or from splitting CO₂ and subsequent burial of reduced carbon. Both are possible on a planet with oxygenic photosynthesis. Most discussions of the history of oxygen on Earth focus on CO₂ and carbon burial because H escape today is negligible.

It was briefly debated in the early 1950s (see Holland, 1962, for a discussion) whether oxygen came from water photolysis followed by hydrogen escape or from photosynthesis followed by incomplete decay of organic matter. This matter was resolved in favor of photosynthesis shortly after the question was asked (Holland, 1962, 1964), but the clear bifurcation in thinking remained after the reason for it had gone. It is to this day implicitly assumed that hydrogen escape is in a one-to-one correspondence to abiotic water photolysis, while carbon burial is in a one-to-one correspondence to biotic water photolysis. The dichotomy is false, as is obvious from the way we have stated it: it is water photolysis in either case. The sum of all water photolyses can be balanced by the sum of hydrogen escape plus incomplete decay of organic matter. To these oxidation and reduction of iron and sulfur must be added.

There are now 0.37 × 10²⁰ mol of O₂ in the atmosphere. The current rate of carbon burial is estimated to be on the order of 1 × 10¹³ mol/yr (Holland, 2002). At this rate it would take only 4 Myrs to build up the O₂ in the modern atmosphere. The equivalent to another ~5 × 10²⁰ mol of O₂ is stored on the continents as sulfates and sedimentary ferric iron (e.g., banded iron formations); these are obvious products of oxic weathering. It would take only 50 Myrs at current rates to build up these inventories.

The sedimentary reservoir of reduced carbon is bigger than the sedimentary reservoirs of oxidized iron and sulfur. A net accumulation of reduced materials in the crust is to be expected if the surface chemistry is dominated by input of reduced gases from the mantle and hydrogen escape is negligible. This is the relationship to be expected if carbon burial is the reason the surface is oxidized. The ~12 × 10²⁰ mol of reduced carbon in the continents (Claire et al., 2006) correspond to ~120 Myr of accumulation. Evidently oxic weathering of old continental carbon to CO₂ is a major part of the modern oxygen cycle (Holland, 1984). The final victory of O₂ has sometimes been identified with a pronounced excursion of δ¹³C in carbonates to high positive values ca. 2.22–2.06 Ga, called the Lomagundi event (Karhu and Holland, 1996; Bekker et al., 2006; Melezhik et al., 2007; Bekker and Holland, 2012). The isotopic excursion suggests a simultaneous massive burial of reduced carbon. Holland (2009) presents a strawman redox budget for the event, but he also emphasizes that Lomagundi was likely a system out of balance. The ~120 Myr time scale for carbon to accumulate in continents corresponds to the duration of the Lomagundi event.

Two bigger reservoirs of oxidized material in the crust are often neglected. Iron in continental basalts is generally more oxidized than the iron in freshly erupted basalts. The difference suggests that ~20 × 10²⁰ mol of O₂ have gone into oxidizing continental basaltic sulfates and sedimentary ferric iron (Lécuyer and Ricardo, 1999). This is bigger than the sedimentary buried carbon reservoir, which means that the continental crust as well as the atmosphere are oxidized. This is the relationship to be expected if hydrogen escape is the reason the surface is oxidized. The detailed mechanism of iron oxidation may be abiotic (e.g., 3FeO + H₂O → Fe₂O₃ + H₂), but even so the result would be a net source of H₂ to the atmosphere and biosphere while the continents accumulated. If this H₂ passed directly to space untouched by life it might be possible to keep continual oxidation separated from carbon burial — which would leave the hypothesis that O₂ originates by burying carbon abstracted from CO₂ unchanged — but in reality the H₂ would be utilized biologically, thus entangling continental oxidation with the carbon cycle.

The other big crustal reservoir is CO₂ in carbonates. It may or may not be relevant to this discussion. Although CO₂ can be abundant in comets, most carbon in asteroids, comets, and meteorites is reduced. Earth probably accreted much of its carbon in a reduced form. Substantial amounts of reduced carbon in a reduced early mantle cannot be ruled out (Hirschmann and Ddagupta, 2009). Holland (2009) emphasizes that there is little evidence that much of the ~50 × 10²⁰ mol of CO₂ now in the continents was present in them in the Archean; this observation plays an important role in his proposal. Because carbon can be oxidized by ferric iron in a QM mantle, carbon oxidation is not necessarily derivative of H escape; it could also be derivative of high pressure disproportionation of ferrous iron followed by export of metallic iron to the core (McClellan, 2005).

The crustal reservoirs are small compared to what could be lost in the mantle, and it is possible that too much has been made of them. Evidence that the upper mantle has not changed its oxidation state since the Hadean (Delano, 2001; Canil, 2002; Trail et al., 2011) tells us little about the lower mantle. The effect of export to the mantle could go either way. Oxidation of the surface could have been accomplished by net export of reduced carbon (Hayes and Waldbauer, 2006) or reduced sulfur rather than by H escape. On the other hand oxidized iron could have been exported to the mantle. The latter might have been especially
efficient in the Archenian for iron-rich banded iron formations, as these can be dense enough to sink into the mantle without the aid of plate tectonics.

5. When did oxygen first appear?

The geological record of the anoxic-oxic transition between ca. 2.4 Ga and 2.0 Ga is often referred to as the “Great Oxidation Event,” usually written “GOE.” Authorities differ on what the GOE means, on how broadly it should be defined, on what events should be included in it, or on what geologic indicators should take precedence (in keeping with the theme of this issue, we have used one of Holland’s definitions). Fig. 1 provides a cartoon summary of some key events. The classical evidence that is concisely summarized by Holland (1999). Holland (2006) provides a more recent summary that includes the post-classical evidence found in the isotopes of sulfur. Changes that mark the GOE include the near disappearance of reduced detrital minerals such as sulfides, siderite and uraninite from streambeds (Rasmussen and Buick, 1999; Hofmann et al., 2009); changes in the weathering of redox sensitive continental elements (Anbar et al., 2007; Frei et al., 2009; Reinhard et al., 2009); the appearance of oxidized surfaces (redbeds) in the place of unoxidized surfaces of comparable sediments (graybeds; Holland, 1999); the first appearance of massive sulfate deposits (Melezhik et al., 2005); the changing redox state of soils revealed by redox sensitive elements such as cerium and iron; the evolution and eventual disappearance of massive banded iron formations (this was drawn out well into Proterozoic; Canfield, 2005); fluctuations in iron isotopes that speak of biological meddling (Rouxel et al., 2005); and looming above all, the abrupt end of huge thick and widespread black shales ca. 3.2 Ga as suggestive of aerobic photosynthesis. The Pongola ice ages ca. 2.94 Ga (Young et al., 1998) could signal the disappearance of major greenhouse gases, probably CO₂ or CH₄ or both, or even H₂ (Wordsworth and Pierrehumbert, 2013). Cold climates can be interpreted as weak circumstantial evidence for oxygen, the enemy of CH₄ and H₂ (but not of CO₂). On the other hand, the Pongola climate could be explained by an anti-greenhouse organic haze (Domagal-Goldman et al., 2008; Zahnle, 2008) triggered by an increase of biogenic CH₄. The massive Steep Rock carbonate reefs suggest a great leap forward in biological productivity, as if some limit had been breached (Wilks and Nisbet, 1988). They are older than 2.78 Ga and probably younger than 2.82 Ga (Fralick et al., 2008).

Next oldest are molecular fossils recovered from ca. 2.78 Ga rocks indicative of sterol synthesis (Brocks et al., 2003; Eigenbrode et al., 2009). Biological synthesis of the original molecules demands the presence of O₂ itself (Summons et al., 2006; Waldbauer et al., 2011), which would leave little room for quibbling, were there not significant doubts that the molecular fossils are as old as the rocks (Rasmussen et al., 2008).

The earliest evidence that water was used as the hydrogen source in photosynthesis is found at 2.72 Ga in lacustrine stromatolites. The Tumbiana stromatolites are biologically, biogenic, and “evidently phototrophic” (Buick, 2008). There is no evidence of any electron donor for photosynthesis apart from water (Buick, 2008). Although stealthy anoxygenic photosynthesis based on H₂ cannot be ruled out (Kharecha et al., 2009), evidence of bubbles in the stromatolites, plausibly interpreted as bubbles of O₂ (Bosak et al., 2009), and the presence of structural features that are claimed to be characteristic of cyanobacterial mats (Flannery and Walter, 2012), strengthen the case for oxygenic photosynthesis. Extremely low δ¹³C suggests both methanogenesis and methanotrophy (Hayes, 1994). Methanotrophy suggests that O₂ was consumed metabolically, as anaerobic methanotrophy using sulfate would seem unlikely given the dearth of sulfur in these rocks. A similar but more widespread prevalence of low δ¹³C in deepwater carbonates paired with modern photosynthetic δ¹³C in shallow water suggests a more far-reaching impact of aerobic ecosystems between 2.45 and 2.7 Ga, perhaps achieving global reach by 2.45 Ga (Eigenbrode and Freeman, 2006).

Between ~2.8 Ga and 2.45 Ga there are several reports of otherwise insoluble elements and minerals such as molybdenum and rhenium (Anbar et al., 2007), chromium (Frei et al., 2009), and pyrite (Stüeken et al., 2012) being weathered from continents and washed into the sea. These elements are more soluble when more highly oxidized. It is interesting that the signature is seen in Mo and Re but not in U, a pattern that strongly suggests that the source of Mo and Re was oxidative weathering of sulfide minerals in particular (Buick, 2008). These occurrences have been given the memorable name “whiffs of oxygen” (Anbar et al., 2007). However, the reported signals, both elemental abundances and isotopic anomalies, are small compared to more modern, undoubted products of weathering under an O₂-rich atmosphere (Scott et al., 2008). Oxidation of Cr to the Cr⁶⁺ of soluble chromate requires O₂ (Frei et al., 2009). Konhauser et al. (2011) suggest instead that strongly acidic conditions would suffice to mobilize chromium as Cr⁷⁺. Acid indirectly indicates oxidation, as the plausible source of acid is oxidation of iron, sulfur or sulfide, but it may not mean O₂, although Konhauser et al. (2011) prefer both biology and O₂, as today at Rio Tinto. Whatever the details, it seems clear that by 2.6 Ga at the latest H₂O was being used as a hydrogen source in photosynthesis in settings by the sea and that the oxidized product was mobile and locally or regionally abundant. Although O₂ need not have been the product, toxic plumes of biogenic O₂ might account for things nicely.

Redbeds are probably the highest fidelity and most directly pertinent of the classic geological indicators of the atmosphere’s changing redox state (Holland, 2006; Kump, 2008). Redbeds indicate an oxidizing atmosphere at the surface but, as we have seen with Mars, not necessarily a lot of free O₂. Fig. 1 places the graybed–redbed transition at 2.32 Ga and after the second of the three Paleoproterozoic ice ages that began ca. 2.45 Ga (Papineau et al., 2007; Guo et al., 2009).
The latest date that has been considered for the GOE, ca. 2.22 Ga, refers to the onset of the Lomagundi $\delta^{13}$C carbonate isotope excursion (Karhu and Holland, 1996; Bekker et al., 2006; Melezhik et al., 2007) and to the massive Hotazel deposit that follows the last of the great Paleoproterozoic ice ages. This — the last possible moment — is where Kopp et al. (2005) place the origin of oxygenic photosynthesis. This is also around the time that the first massive sulfate deposits appear (Melezhik et al., 2005; Holland, 2006; Schröder et al., 2008). The Lomagundi isotope excursion suggests that a great deal of isotopically light reduced carbon was buried, which on the modern Earth would imply that a great deal of $\text{O}_2$ was left behind in the atmosphere, in obedience to the net reaction $\text{CO}_2 \rightarrow \text{C} + \text{O}_2$. Kopp et al. (2005) suggest that the third ice age was the biggest and therefore the best ice age to link to the origin of oxygen photosynthesis, but this is a weak argument given that at least two great ice ages had come and gone over the previous 200 Myr.

If there is one date to divide the Archean from the Proterozoic, the best choice would be the end of an extraordinary isotopic fractionation history of sedimentary sulfur, which was followed by Earth’s plunge into the first of a series of stupendous ice ages. On the other hand, there remains some uncertainty regarding when these events took place. We have placed it at 2.45 Ga, which is the earliest plausible date for the event and the one most often cited, but a date as late as 2.32 Ga is possible (Bekker et al., 2004; Bekker and Holland, 2012). Both isotopes and climate signal major changes in the composition of the atmosphere (Zahnle et al., 2006).

Sulfur is particularly important because it is, with carbon and iron, one of the three abundant elements that easily change oxidation state at ordinary conditions. The record of sulfur’s ordinary mass fractionation (denoted $\delta^{34}$S) is consistent with increasing oxidation of the surface beginning in the late Archean (Fig. 2). The widening envelope bounding the scattered data indicates that greater amounts of sulfur were becoming biologically available, presumably as soluble sulfate in the seas, which afforded biology the opportunity to be more choosy about which sulfur isotopes they used (Canfield, 2005). The result is a progressively increasing scatter in $\delta^{34}$S from 2.7 Ga to 2.0 Ga.

Fig. 2. Top. Carbon isotope history in carbonates ($\delta^{13}$C ≈ 0) and in reduced carbon ($\delta^{13}$C ≈ −4%). Paleoproterozoic ice ages are indicated with blue shading. Bottom. The widening dispersion of ordinary mass-dependent fractionation $\delta^{34}$S (Canfield, 2005) as a function of time on Earth suggests that sulfur becomes progressively more abundant and less dear to life. Middle. Mass independent S fractionation $\Delta^{33}$S (Farquhar and Johnston, 2008) appears to increase as the biologically accessible sulfur pool grows until the signal abruptly vanishes 2.46 Ga, to be followed immediately by a series of major ice ages. Orange shading in the late Archean marks the time when biogenic reduced gases such as $\text{CH}_4$ and $\text{H}_2\text{S}$ were abundant and biogenic $\text{O}_2$ suppressed.
The extraordinary fractionations that highlight the date of 2.45 Ga are recorded with sulfur’s third isotope. This fractionation, denoted $\Delta^{33}S$, (often called mass independent fractionation of sulfur, or MIF-S for short) is the difference between the fractionation of $^{33}S$ expected from $^{32}S$ and what is actually observed. Large nonzero values of $\Delta^{33}S$ vanish abruptly at ~2.45 Ga, although weak signals somewhat comparable to what can be found on Earth today in Antarctica continue for another ~70 Myrs (Farquhar and Johnston, 2008).

The only known way to get MIF-S big enough to give the observed $\Delta^{33}S$ signal involves atmospheric photochemistry driven by solar UV (Farquhar et al., 2000). An atmospheric source has two requirements. First, the UV cannot be absorbed by ozone. Thus there cannot be much O$_2$ in the atmosphere. Second, there have to be at least two channels to remove sulfur from the atmosphere (Pavol and Kasting, 2002), in order to keep the two complementary $\Delta^{33}S$ signals separate (if, as today, nearly all S left the atmosphere as sulfate aerosols, the $\Delta^{33}S$ would sum to zero). Laboratory experiments (Farquhar et al., 2001; DeWitt et al., 2010) show and photochemical models predict (Pavol and Kasting, 2002; Zahnle et al., 2006) that both sulfate aerosols and elemental sulfur (S$_8$) aerosols can be generated photochemically in an anoxic atmosphere, but that elemental sulfur cannot form if the atmosphere is not sufficiently reduced. Hence significant nonzero $\Delta^{33}S$ demands an effectively anoxic atmosphere ($p_{O_2} \sim 10^{-7}$ bars; Zahnle et al., 2006) before 2.45 Ga. Getting $\Delta^{33}S$ into the sedimentary record raises additional issues in the presence of oceans and biology. Separation is maintained at first because S$_8$ is insoluble in water, but if oxidized or reduced — either biologically or biologically — its products need not be. Halevy et al. (2010) consider the simple but extreme expedient of re-imprinting, and potentially amplifying through distillation, the dust, while soluble sulfate is washed into the sea. Another factor to consider is that on average about 20% of the CO$_2$ outgassed from volcanoes were necessarily consistently appreciable at the surface (Zahnle et al., 2011). One may protest that surely the Archean was more volcanically active than today, and as volcanic gases are reducing, the Archean atmosphere should have been more reduced than today, but by itself this gains us nothing if the ratios of H$_2$ and CO to CO$_2$ in volcanic gases did not change: in this theory, it is the ratio of H$_2$ to CO to CO$_2$ in volcanic gases that is important. Simply changing the rate of volcanic degassing while holding the composition of the gases fixed does not work, because the burial rates of C and carbonate (CO$_2$) must both keep pace with the volcanic supply of CO$_2$ if $\delta^{13}C$ be constant. Some other change is required.

The simplest change is to postulate a higher H$_2$/CO$_2$ ratio in Archean volcanic gases (Holland, 2009). This would give a higher H$_2$/CO$_2$ at a fixed $f_{O_2}$ and thus leave more reducing power in the atmosphere. A change in the H$_2$/CO$_2$ ratio coming out of volcanoes implies a proportionate change in the H$_2$O/CO$_2$ ratio in the rocks that go into the volcanoes. The evidence for this is massive carbonates being rare in continental sediments before 2.8 Ga (Holland, 2009). Against this is the expectation (prejudice) that, as the mantle cools, it will degas less CO$_2$ and incorporate a greater fraction of the carbonate and organic carbon entering subduction zones, which makes a true dearth of CO$_2$ in the Archean a bit of a surprise. Holland (2009) shows that it is possible to get the sought-for behavior from a model that divides volcanic CO$_2$ emissions into a declining juvenile component and a growing component stemming from recycled crust.

A bigger effect of the same sort can be obtained from sulfur gases, which are now a major player in the redox budget of volcanic gases, but may not have been if sulfur were scarce in subducting zones. Holland (2002) developed a model that left H$_2$/H$_2$O fixed but added progressively increasing amounts of sulfur gases to the system. The evolution of the envelope of $\delta^{34}S$ in the sedimentary record suggests that sulfur was scarce before 2.7 Ga and became abundant by 2.0 Ga, and thus that its role grew from a trace element to a major player in the redox budget at roughly the same time that $O_2$ took over. In both the 2002 and 2009 models, Holland accepts a constant mantle $f_{O_2}$ and fixes the carbon burial rate to 20% of the volcanic CO$_2$ flux (i.e., $f_{org}$, the fraction of volcanic CO$_2$ buried as C, is 0.2). Adding sulfur creates a more complicated relation between volcanic fluxes and carbon burial, the gist of which is that, while volcanoes mostly emit S as SO$_2$, Holland (2002) argues that S is removed from the anoxic atmosphere mostly as FeS$_2$. Thus volcanic H$_2$ and CO are tasked with reducing SO$_2$ to FeS$_2$ in addition to reducing the $O_2$ that comes from carbon burial and H escape. As the amount of sulfur in the crustal rock cycle increased with time, the amount of S-containing volcanic gases increased, and so the burden of reducing SO$_2$ increased until it overcame the ability of H$_2$ and CO to reduce it all. At this point $O_2$ wins. Holland (2009) later modified the argument to allow H$_2$O/CO$_2$ to decrease as well, which enhances the effect.

Holland’s models have many points to recommend them, but they cannot be the whole story because they leave the rough constancy of $f_{org}$ unexplained, they do not explain why carbonate or sulfate accumulate at the surface, and they do not explain why the surface began to
oxidized. The models resemble a titration, with sulfur and carbonate being added to the volcanoes until the reducing power (volcanic H2 and CO vs. C burial) is overcome. Indeed, Holland’s models fit rather well with hydrogen escape as the driving principle.

7. Hydrogen escape

Hydrogen escape oxidizes planets, beginning at the surface. Earth is no exception. The pertinent questions are by how much and at what rate. Our hypothesis is quite simple: hydrogen escape oxidizes the surface and atmosphere until they become oxidized enough that O2 can accumulate (Catling et al., 2001, 2005; Claire et al., 2006). It applies directly and clearly to a model like Holland (2002), where the key oxidized material to be accumulated is sulfate (Zahnle et al., 2006). It could be applied to Guo et al’s (2009) suggestion that oceanic Fe3+2 was titrated between 2.45 Ga and 2.32 Ga. It can also be applied to the oxidation of continental igneous rocks, which as noted above appear to hold the largest reservoir of oxidized material at the surface (Lécuyer and Ricard, 1999; Claire et al., 2006). In all cases what results is a kind of oxidation hourglass.

To quantify things we can ask how long it takes for plausible levels of hydrogen escape to oxidize representative crustal inventories. The rate at which hydrogen escapes from Earth is well approximated by Hunter’s diffusion limit (Hunten and Donahue, 1976), which can be evaluated as

\[ \phi_{\text{H}_2} = 5.7 \times 10^{15} f_{\text{tot}}(\text{H}_2) \text{mol yr}^{-1}, \]

where \( f_{\text{tot}}(\text{H}_2) = f(\text{H}_2) + f(\text{H}_2O) + 2f(\text{CH}_4) \) ... is the total mixing ratio of H2 equivalents in the stratosphere. In the current atmosphere the effective H2 mixing ratio \( f_{\text{H}_2} \) is 7.1 ppmv, to which CH4 and H2O contribute equally (Nassar et al., 2005). Diffusion limited escape today corresponds to the loss of 1 m of water per billion years, or equivalently the production of 0.1 bar of O2 per billion years.

By contrast, hydrogen escape rates in the anoxic Archean atmosphere were probably not negligible (Hunten and Donahue, 1976). Through methane, hydrogen escape is at least in part under biological control, even today (Hunten and Donahue, 1976). The modern-day biogenic CH4 flux vented into an anoxic atmosphere would support a methane mixing ratio of 1000 ppmv (Kharecha et al., 2005; Haqq-Misra et al., 2008). In the diffusion limit this corresponds to an H2 escape rate of 1015 mol/yr. This corresponds in turn to a crustal oxidation rate half that generated today by carbon burial. Other ways to express this rate are as the loss of 400 m of water over 1 Gyr, or as the creation of 40 bars of O2 over 1 Gyr. This is enough bleeding power to oxidize the continents (basalts + sediments — buried carbon) in some 300 Myr. The biogenic methane flux and the consequent oxidation rates may have been bigger than this following the advent of oxidative photosynthesis, but the present biogenic methane flux would be big enough to do the job.

Holland was no enemy of hydrogen escape as a possible major player in the oxidation of the Earth, as he makes clear in his books The Chemistry of the Atmosphere and Oceans (Holland, 1978) and The Chemical Evolution of the Atmosphere and Oceans (Holland, 1984). Holland (1978, p. 296) wrote, “Hydrogen escape from an anoxic atmosphere containing sizable quantities of reducing gases is probably orders of magnitude more rapid than it is from the atmosphere today, and could have had an important influence on the chemistry of the early atmosphere of the Earth.” He always included hydrogen escape in his budgets (as for example in Holland, 2009), but he usually also pointed out that the term was small compared to the other terms in the redox budget. As a global oxidizing force he appears to have considered it only in relation to the mantle’s volcanic H2 emissions and the resulting drift of the mantle’s oxygen fugacity. When the latter was found not to have changed over the past 3.8 Gyr, he appears to have given up on hydrogen escape.

But judging the importance of hydrogen escape solely by the magnitude of the term misses the greater point that hydrogen escape is irreversible. Carbon buried in the continents will over the course of the rock cycle be lifted up, exposed to weathering, consume its equivalent of oxygen, and be reborn as CO2. The oxidation generated by carbon burial is provisional. The oxidation generated by hydrogen escape is permanent. This is a qualitative distinction that makes it misleading to compare the magnitude of the hydrogen escape flux against the other major terms in Earth’s redox budget. In a sense, it can only be compared to the oxygen escape flux, which is very nearly zero.

8. Discussion

If the surface is to oxidize as the influence of the mantle wanes, there must exist some innate tendency of the surface to oxidize. Hydrogen escape is such a tendency. The process is general and should apply to all habitable planets of all sizes and types throughout the universe. It is therefore testable by future astronomical observations.

The fundamental importance of hydrogen escape is not a position that one often encounters in Earth science. The most common argument is that oxygen is a secondary consequence of Earth cooling. Arguments derived from Holland’s (1962) hypothesis usually compare the flux of reducing power from the mantle to a fixed flux of reduced matter that accumulates on continents or is buried in sediments that are destined to be subducted by the mantle. The reduced matter can be carbon made from CO2 or metal sulfides made from volcanic SO2. The implicit presumption is that production and burial of reduced matter can be regarded as constant rates that are independent of the redox state of the surface environment. The constancy may be enforced by biological activity. If the mantle supplies more reducing power than is needed to meet the constant carbon burial rate, the waste hydrogen is vented to space. If the mantle supplies less reducing power than needed for the fixed burial rate, the surface and atmosphere become oxidized. If by this time oxygenic photosynthesis had been invented, the atmosphere would then begin to accumulate O2. The redox budget can be balanced by such a model, but one is ultimately left with the hypothesis that an oxidized surface environment oxidizes itself by reducing an already reduced mantle.

In most respects secular cooling is a more important influence on planetary evolution than is hydrogen escape. But cooling does not in itself alter oxidation. Oxygenation by cooling would have to be a by-product, a secondary effect that preferentially pools oxidants at the surface. Speculations abound. In addition to those which closely follow Holland’s, which we have discussed, a nonexhaustive list includes (i) declining rates of volcanism (a fact that does not explain why oxygen pools at the surface); (ii) a switch from submarine to subaerial volcanism that diminishes the reducing power of volcanic gases (Kump and Barley, 2007; Gaillard et al., 2011); (iii) increased stability and volume of continents as burial grounds of reduced (or oxidized) materials; (iv) a switch from weathering of more mafic to less mafic matter, driven by the growth of continents, that changes the relative burial rates of reduced and oxidized minerals (Slep, 2005); (v) increased supply of P (as limiting nutrient) from continental weathering consequent to continental growth; (vi) less or more subduction; (vii) increased stability of relatively volatile reduced (or oxidized) minerals during subduction (Hayes and Waldbauer, 2006, in effect); (viii) and internal differentiation within the mantle (Kump et al., 2001; McCammon, 2005). Several of these change the magnitude but not the sign of surface oxidation; several of these include a measure of knowing in advance what the answer needs to be.

Three other long term trends should be mentioned. (i) The Sun grows monotonically brighter. How this would impact oxygen is unclear, but it could do so indirectly if the climate were moderated by greenhouse gas feedbacks (in which O2 can play a part). Such an argument is usually associated with biological meddling (Goldblatt et al., 2006; Zahnle, 2008). (ii) The flux of high energy ionizing radiation monotonically decreases. The effect is mostly on hydrogen escape, which would be easier in the past than it is today. (iii) Biological evolution
might be innately progressive, with a trend to develop more potent metabolicisms. Free oxygen is an outstandingly potent and useful fuel for animal metabolism (Catling et al., 2005). The hypothesis that biological evolution is destined to make oxygen is at present best regarded as speculative.

Here we argue that oxidation and oxygen are consequences of hydrogen escape, which is the one process affecting planetary oxidation that always points in the same direction. We do not argue that hydrogen escape acts alone without interaction with biological innovation or planetary evolution. We do argue that hydrogen escape determines the direction of things, especially at the planet's surface where its effects are most immediate. This includes the speculation that hydrogen escape provides the bias to biological evolution. We suggest that the hundreds of millions of years that elapsed between the invention of oxygenic photosynthesis and the establishment of an oxygen atmosphere were how long it took to oxidize the continents by hydrogen escape. This is complicated by, and probably linked to, the putative growth of continents between 3 Ga and 2 Ga. Growing continents allow more reduced carbon, sulfate, BIF, and carbonate to be stored on continents, and increase the amount of oxidized igneous rock in the cratons.

It has been argued that continental oxidation as registered in basaltic has nothing to do with oxygen, because basalt is oxidized abiotically shortly after eruption by water, not by oxygen or sulfate or any other biogenic oxidant (Kasting, 2001). However, H₂ generated when basalts are oxidized by H₂O is utilized biologically to make organic matter from CO₂. Hence the apparent net oxidation of continents kept pace with hydrogen escape. This is complicated by, and probably linked to, the putative growth of continents between 3 Ga and 2 Ga. Growing continents allow more reduced carbon, sulfate, BIF, and carbonate to be stored on continents, and increase the amount of oxidized igneous rock in the cratons.

It is intriguing to turn all this on its head and ask whether continental oxidation and continental growth are both governed by hydrogen escape — i.e., to ask whether continents as we know them must be relatively oxidized to exist as such, and therefore that hydrogen escape determines the rate of continental growth (cf. Jagoutz, 2013). But whatever the details, oxygen’s conquest waited on the exhaustion of the powers of reduction. Hydrogen escape was Earth’s hourglass.

Acknowledgments

The authors thank A. Bekker, R. Buick, J. Farquhar, J. Kasting, L. Kump, and E. Strucl for insightful reviews, data, or both. The authors thank the NASA Exobiology Program and the NASA National Astrobiology Institute for support of this work. DCC acknowledges support from the NASA Exobiology grant number NNX10AQ90G.

References


