

# Toward a List of Molecules as Potential Biosignature Gases for the Search for Life on Exoplanets and Applications to Terrestrial Biochemistry

S. Seager,<sup>1,2</sup> W. Bains,<sup>1,3</sup> and J.J. Petkowski<sup>1</sup>

## Abstract

Thousands of exoplanets are known to orbit nearby stars. Plans for the next generation of space-based and ground-based telescopes are fueling the anticipation that a precious few habitable planets can be identified in the coming decade. Even more highly anticipated is the chance to find signs of life on these habitable planets by way of biosignature gases. But which gases should we search for? Although a few biosignature gases are prominent in Earth's atmospheric spectrum (O<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O), others have been considered as being produced at or able to accumulate to higher levels on exo-Earths (*e.g.*, dimethyl sulfide and CH<sub>3</sub>Cl). Life on Earth produces thousands of different gases (although most in very small quantities). Some might be produced and/or accumulate in an exo-Earth atmosphere to high levels, depending on the exo-Earth ecology and surface and atmospheric chemistry.

To maximize our chances of recognizing biosignature gases, we promote the concept that all stable and potentially volatile molecules should initially be considered as viable biosignature gases. We present a new approach to the subject of biosignature gases by systematically constructing lists of volatile molecules in different categories. An exhaustive list up to six non-H atoms is presented, totaling about 14,000 molecules. About 2500 of these are CNOPSH compounds. An approach for extending the list to larger molecules is described. We further show that about one-fourth of CNOPSH molecules (again, up to  $N=6$  non-H atoms) are known to be produced by life on Earth. The list can be used to study classes of chemicals that might be potential biosignature gases, considering their accumulation and possible false positives on exoplanets with atmospheres and surface environments different from Earth's. The list can also be used for terrestrial biochemistry applications, some examples of which are provided. We provide an online community usage database to serve as a registry for volatile molecules including biogenic compounds. Key Words: Astrobiology—Atmospheric gases—Biosignatures—Exoplanets. *Astrobiology* 16, 465–485.

## 1. Introduction: Motivation

**T**HE SEARCH FOR biosignature gases in the atmospheres of exo-Earths as a sign of life is a guiding goal for many in the exoplanet community. Biosignature gases are gases produced by life that accumulate in the atmosphere and could be detectable remotely by space telescopes. Given the observational evidence of the vast diversity of exoplanetary physical properties and the wide variety of the gases produced

by life on Earth, a new approach in the study biosignature gases is warranted.

### 1.1. A brief background to biosignature gases

For more than half a century, researchers have considered the possibility of inferring the presence of life on planets other than Earth. In fact, the concept of oxygen as an atmospheric biosignature gas was mentioned over 80

<sup>1</sup>Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts.

<sup>2</sup>Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts.

<sup>3</sup>Rufus Scientific, Cambridge, UK.

years ago<sup>1</sup> (Jeans, 1930). Early work that remains part of the paradigm for exoplanet life-detection focused on gases severely out of thermodynamic equilibrium, arguing that only life could maintain thermodynamic disequilibrium in the atmosphere. Specifically, the detection of oxygen (O<sub>2</sub>) and methane (CH<sub>4</sub>) (Lederberg, 1965; Lovelock, 1965) was suggested as the most robust atmospheric evidence that Earth supported life. O<sub>2</sub> and its photolytic product ozone (O<sub>3</sub>) as well as CH<sub>4</sub> have been extensively studied as biosignature gases in their own right (*e.g.*, Léger *et al.*, 1996; Schindler and Kasting, 2000; Des Marais *et al.*, 2002; Segura *et al.*, 2003; Kaltenegger *et al.*, 2007). The history of O<sub>2</sub> as a biosignature gas and the pros and cons of a thermodynamic disequilibrium as a sign of life are critically reviewed in Seager and Bains (2015). The subject of O<sub>2</sub> false positives has been considered with growing interest (Schindler and Kasting, 2000; Selsis *et al.*, 2002; Léger *et al.*, 2011; Hu *et al.*, 2012; Domagal-Goldman *et al.*, 2014; Tian *et al.*, 2014; Wordsworth and Pierrehumbert, 2014; Harman *et al.*, 2015; Luger and Barnes, 2015). Part of the exoplanet community hopes that future-generation telescopes will obtain high-enough quality spectra to provide the planetary environmental context with which to enable sufficient confidence in the identification of O<sub>2</sub> as a biosignature gas.

Gases (other than O<sub>2</sub> or CH<sub>4</sub>) that are produced by life on Earth have also been studied, including nitrous oxide (N<sub>2</sub>O) (Des Marais *et al.*, 2002), dimethylsulfide (DMDS) (Pilcher, 2003), methyl chloride (CH<sub>3</sub>Cl) (Segura *et al.*, 2005), and dimethyl sulfide (DMS) and other sulfur gases (Domagal-Goldman *et al.*, 2011). The planetary environment influences the destruction of gases, including surface chemistry and especially the host star EUV radiation that drives photochemistry in the planetary atmosphere (*e.g.*, Segura *et al.*, 2003; Tian *et al.*, 2014). Yet other than differing stellar radiation environments, most biosignature gas research to date focuses on Earth-like planets—that is, planets with Earth-like atmospheres and Earth-like bioflux gas sources.

The notion of habitability is anchored in the concept of surface temperature suitable for life, specifically for the existence of liquid water on the surface. Because the planetary atmosphere masses and compositions—and hence a planet's greenhouse effect that controls the surface temperature—are unknown and expected to be varied (Elkins-Tanton and Seager, 2008), habitability is planet-specific (Seager, 2013). A wide diversity of planet types are expected, with all masses, sizes, and orbits apparently in existence, as far as observations can ascertain (Howard, 2013; Winn and Fabrycky, 2015). Planets orbiting well interior (Abe *et al.*, 2011; Zsom *et al.*, 2013) or exterior (Pierrehumbert and Gaidos, 2011) to an Earth-like planet's habitable zone boundaries (Kopparapu *et al.*, 2013) must be considered. The planet di-

versity also could well extend to the surface sources and sinks of gases, especially the redox state of the planetary surface—a wide variety of habitable planets have been hypothesized in this regard, from water worlds (Kuchner, 2003; Léger *et al.*, 2004), to planets with hydrogen-rich atmospheres (Pierrehumbert and Gaidos, 2011; Seager *et al.*, 2013a), to Venus-like worlds (Schaefer and Fegley, 2011) and planets with increased volcanism (Kaltenegger and Sasselov, 2010; Hu *et al.*, 2013). The extreme and far UV radiation that drives atmospheric photochemistry will vary depending on host star type and age (Guinan *et al.*, 2003; Shkolnik and Barman, 2014). Last but not least, there should be a diversity of net bioflux emission levels on an exoplanet, but any estimation is beyond current solution (Seager *et al.*, 2013b). This physical and chemical diversity will affect which molecules accumulate in a planet's atmosphere and what false-positive gas sources might occur. The diversity of known exoplanets and the anticipation of diversity of habitable planets therefore further motivate our investigation of alternative volatile biomolecules that might be significant in the atmosphere of a planet other than Earth.

## 1.2. Gases produced by life (on Earth)

In addition to the diversity of exoplanets, we must recognize the diversity of molecules produced by life on Earth.

The chemicals produced by life on Earth are numbering in the hundreds of thousands [estimated from plant natural products (Gunatilaka, 2012), microbial natural products (Sanchez *et al.*, 2012), and marine natural products (Fusetani, 2012)]. But only a subset of hundreds of these are volatile enough to enter Earth's atmosphere at more than trace concentrations. Only a few of the gases produced by life on Earth—O<sub>2</sub> (and O<sub>3</sub>), CH<sub>4</sub>, and N<sub>2</sub>O—have been detected in Earthshine and spacecraft observations of the spatially unresolved “Earth as an exoplanet” (*e.g.*, Christensen and Pearl, 1997; Turnbull *et al.*, 2006; Palle *et al.*, 2009; Robinson *et al.*, 2011). These observations show us what might be possible from space telescopes capable of finding and characterizing Earth-like planets orbiting Sun-like stars in reflected light observations (*e.g.*, Seager *et al.*, 2015; Stapelfeldt *et al.*, 2015).

It is interesting to recognize that life produces all the gases in Earth's atmosphere (specifically the troposphere) present at the parts-per-trillion level by volume or higher (see Appendix Table A1), with the exception of the noble gases. Out of 47 known volatiles in Earth's atmosphere observed at the parts-per-trillion level, 42 are known to be biogenic, though not exclusively so. If we exclude the 10 entirely anthropogenically produced chlorofluorocarbons, the numbers reduce to 37 and 32, respectively. Most of Earth's atmospheric gases are, of course, not unique to life; moreover, life is not the dominant source of atmospheric gases for most cases. Some atmospheric gases are already a basic atmospheric constituent (*e.g.*, N<sub>2</sub>, CO<sub>2</sub>, and H<sub>2</sub>O). Many are produced by geological processes (*e.g.*, CH<sub>4</sub> and H<sub>2</sub>S). However, the relative rate of production of a gas by life is a function of both geological and biological production rates, and both are specific to the planet. On other worlds, biology could be the dominant source of any gas.

The gases produced by life on Earth, when organized into two broad categories, create a conundrum. The gases

<sup>1</sup>“It seems at first somewhat surprising that oxygen figures so largely in the earth's atmosphere, in view of its readiness to enter into chemical combination with other substances. We know, however, that vegetation is continually discharging oxygen into the atmosphere, and it has often been suggested that the oxygen of the earth's atmosphere may be mainly or entirely of vegetable origin. If so, the presence or absence of oxygen in the atmosphere of other planets should shew whether vegetation similar to that we have on earth exists on those planets or not.”

expected to be produced in abundance by life are ones that are also rife with false positives. These are by-product gases from biological energy extraction from chemical potential energy gradients (combinations of chemicals that are out of thermodynamic equilibrium) that are common and geochemically produced. Such chemical potential energy gradients are widely exploited by life on Earth, but geochemistry has the same gases to work with as life does. So while in some environments life is needed to catalyze the reaction of the disequilibrium chemicals, in others the same reactions will be spontaneously occurring. An example is methane. Methane is a by-product of methanogenesis, but it is also released from vents at mid-ocean ridges because of hydrothermal, abiotic chemistry.

The second broad category of gases is the class of biosignature gases produced for secondary or unknown reasons, such as stress or signaling. Such gases are often organism- and mechanism-specific and hence are expected to be produced in small quantities. But because they are so specialized they will in most cases likely not have geochemical sources that could lead to a false positive in the search for life. Some of these specialized gases are produced in amounts sufficient to affect overall atmospheric chemistry (such as isoprene or DMS) and to possibly be detected remotely on exoplanets (such as methyl chloride; Segura *et al.*, 2005). Which gases are produced in large amounts and which are produced only as trace gases is determined by the functional needs of the organism. The maximum productivity is always constrained by resource limitations, diffusion limits, and energy availability. In the absence of knowledge of the function served by a gas in an alien ecology, the choice of which gas is made in large amounts can appear arbitrary.

For completeness, we mention two other classes of biosignature-related gases. One class is biosignature gases produced as by-product gases from energy-requiring metabolic reactions for biomass building. On Earth these are reactions that capture environmental carbon (and to a lesser extent other elements) in biomass. An example is photosynthesis, which produces O<sub>2</sub>. The other category is photochemical or chemical reaction by-products of biosignature gases, such as O<sub>3</sub> as produced from O<sub>2</sub>. For a more detailed discussion of the above-described classes of biosignature gases, see Seager *et al.* (2013b).

We choose to focus on gases emitted by life, rather than solid products or features. For example, while the vegetation “red edge” (for a description, see the review by Arnold, 2008) and other spectral features due to pigments in vegetation or bacteria have been studied as biosignatures in reflected light of Earthshine, their signals are weak (partly due to limited surface coverage) and diminished by clouds (Montañés-Rodríguez *et al.*, 2006). For other planets, surface biosignatures are another area of research (spectra, cover, signal strength) and are not within the scope of this paper. Similarly, technological signatures as signs of life on exoplanets are not within the scope of this paper.

### 1.3. A new approach

All life on Earth makes gas products, and basic chemistry suggests the same will be true of any other plausible biochemistry. The question is, what products? Given the multitude of gases produced by life on Earth and the very different planetary

atmosphere, surface, and stellar radiation environments anticipated on exo-Earths, we are motivated to propose a new approach. The first step is to come up with a list of *all* molecules that are stable and potentially volatile, not just the molecules produced by life on Earth. The second step is to consider each molecular gas and its viability as a biosignature gas on exo-Earths within different exoplanetary atmosphere and surface environments and based on the strength and wavelength range of its spectroscopic signature. This second step must also include an assessment of false positives and whether or not other gases and their spectral features might support a false-positive scenario. The third step is in the future, to use the future space-based telescopes to search for these candidate biosignature gases on yet-to-be-discovered exo-Earths.

This paper describes the first step: generation of a list of molecules for biosignature gases. As an illustration of the utility of the list beyond exoplanet research, we comment on our findings on the fraction of chemicals in the list that are produced by terrestrial life.

## 2. Methods

We construct a list of molecules that are stable and likely volatile as pure compounds at standard temperature and pressure (STP). Materials made of small molecules are likely to be volatile, meaning more likely to be in gas form in a planetary atmosphere. Our list is therefore constructed starting with small molecules. We present a list of molecules up to  $N=6$  non-H atoms. In Section 4 we cover limitations for our method choices, including STP and proxy for volatility, and in addition describe how to extend the list to all molecules that might be volatile and stable.

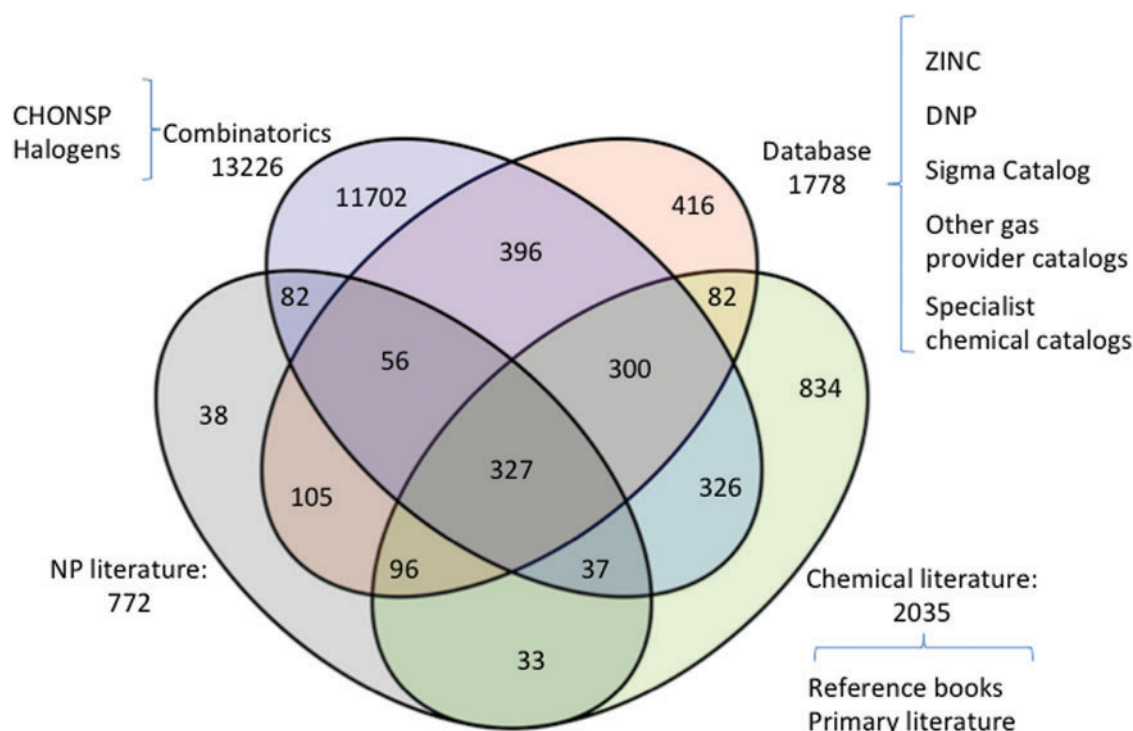
To construct the list we divide molecules into three categories, for practical purposes.

The first category is that of molecules that only contain the elements C, N, O, P, S, and H, “the CNOPSH List” (see Section 2.1). The second category is defined as any molecule not falling into our first category that also does not contain a halogen atom, “the Extended CNOPSH List” (see Section 2.2). The third category is halogenated compounds, which, although they fall into either the first or second categories, are considered on their own due to the extensive numbers of such compounds, “the Halogenated Compounds List” (see Section 2.3). A summary of the approaches that went into construction of the list of small molecules for all three of our categories is illustrated in Fig. 1.

### 2.1. The CNOPSH List

To construct the list of CNOPSH compounds, we took two approaches. The first started with a combinatorics approach, and the second was a supplemental database search (described in Section 2.2.2).

**2.1.1. Combinatorics approach for CNOPSH.** The first approach in the list construction of CNOPSH compounds is to consider molecules built from the six principal biogenic elements, carbon (C), nitrogen (N), oxygen (O), phosphorus (P), sulfur (S) and hydrogen (H). We call this the CNOPSH category of molecules. We wrote a chemical combinatorics algorithm to generate all possible molecular formulae with up to and including six non-H atoms, a cutoff that is explained below.



**FIG. 1.** Illustration of the functional approach to building the list for all small, stable, volatile compounds, for molecules with  $N \leq 6$  non-H atoms. Four types of sources of molecules are shown, one in each segment of the figure. The top left shows the compounds that came out of combinatorial construction (for CNOPSH and halogenated compounds), with the total number listed outside the diagram. The top right shows the molecules from database trawling. The bottom right shows compounds obtained from searching the chemical literature (mostly for the P-block and metal-containing compounds within the Extended CNOPSH List), and the bottom left shows the chemical compounds from a targeted search for chemicals made by life, especially low-molecular-weight compounds. The numbers inside the diagram give the number of molecules found in the respective searches and show the overlap. For example (top right), 416 molecules were found in chemical databases only, and 82 were found in both databases and the chemical literature search.

We describe a molecule by its non-H atoms because it enables a simple implementation of the structural definition of a molecule as a set of non-H atoms connected to each other by at least one bond each. Any valency not involved in CNOPS bonding is assumed to be populated with an H atom. Each CNOPS combination can have a different number of H atoms populating the open valencies, depending on the arrangement and order (*i.e.*, single, double, or triple) of bonds present among the non-H atoms. In other words, several molecular formulae are generated from each CNOPS combination because different bonding patterns are possible between the atoms, with all unfilled valences filled with H atoms. [For example,  $C_2O$  could be  $C_2H_6O$  (ethanol or dimethyl ether), or  $C_2H_4O$  (acetaldehyde or ethylene oxide).] We are interested in volatile molecules, so we emphasize that the atomic weight of H is so relatively small that addition of H atoms does not make a substantial difference to the molecular weight (a physical property related to volatility).

We chose the cutoff value of  $N=6$  non-H atoms for purely pragmatic reasons. The number of possible molecules is an exponential function of the number of non-H atoms (Bains and Seager, 2012). As we had to manually curate all the nonhalogenated molecules for stability and volatility, and manually search the literature for life production, a cutoff of  $N=6$  was imposed to keep the curation practical. The advantage of adopting the  $N=6$  cutoff is that a substantial

fraction of the molecules will be volatile. As the molecules get larger, fewer and fewer will be volatile, so predicting volatility becomes more and more important (see Section 4.1.1).

We could include H atoms in our structural definition and limit ourselves to  $N=6$  atoms including H. However, this would exclude many simple molecules known to be important for life, such as ethane ( $C_2H_6$ ), and would bias the approach toward generation of more oxidized molecules. Alternatively, we could consider all molecules with a large number of atoms (*e.g.*,  $N=12$ ), which would include ethane. However, this would include an unmanageably large number of molecules in total (possibly as many as  $3 \times 10^7$ ; Bains and Seager, 2012).

In summary, while the list description of molecules “with up to  $N=6$  non-H atoms” is an awkward formulation, it is a compromise that avoids artificial biases in generation of our list, includes molecules already studied as biosignature gases, and ensures the practicality of having to research each molecule by manual literature search.

In the combinatorics code, CNOPS atoms were put together in every possible combination. To each of these combinations, the code added the maximum number of H atoms, assuming the lowest-order bonding between atoms, that is, assuming the highest number of valencies were left open for each atom. The computer code then proceeded to consider

lower numbers of available valencies with fewer H atoms added. Numerically, the algorithm for the number of H atoms added can be summarized by  $N_H = NV - 2*(a - 1) - 2*n$ , where  $a$  is the number of non-hydrogen atoms;  $n$  is an integer  $n=0, 1, 2, 3 \dots$  such that  $N \geq 0$ ;  $NV$  is the number of available valence slots for hydrogen atoms (assessed by summing  $C=4, N=3, O=2, P=3, S=2$ ). We note that any missing oxidation states were later included during a chemical databases search approach (see Section 2.2.2).

Not all the molecular formulae will represent realistic chemical structures (*i.e.*, real molecules), and some molecular formulae represent more than one chemical structure. To identify which molecular formulae in the list actually describe one (or more) real chemical structures, we queried the ChemSpider database to identify chemical structures matching each formula. This query process ruled out some number of formulae as not representing real molecules and identified that other formulae represented more than one real molecule. We eliminated radical or molecular fragments from the data set by confirming the reality of molecular structures in the PubChem database, which is focused on experimental study of organic chemicals and hence is more likely to be populated by molecules that actually exist. This filtered out about 40% of the ChemSpider hits (radicals or molecular fragments). After the above filtering, the combinatorically generated CNOPSH list contained about 1300 molecules. After the database and chemical supplier searches (see Section 2.2.2), the number totaled about 2500.

## 2.2. The Extended CNOPSH List

We call our list of compounds beyond strictly CNOPSH the Extended CNOPSH List. This list includes both organics and inorganics.

2.2.1. Manual search for the Extended CNOPSH List. The number of inorganic and organic, stable, volatile molecules that add to the CNOPSH List is much smaller than those in the CNOPSH List itself; hence we proceeded with a manual search. The first approach was to explicitly search the ChemSpider database for any molecule with a P-block element, up to a molecular mass of 400 Da. There is a high probability that anything with a molecular mass larger than 400 Da will not be significantly volatile. More specifically, the search was for compounds likely to be stable to hydrolysis containing at least one atom from the nonradioactive, non-halogen P-block elements other than CNOPS (*i.e.*, B, Al, Ga, In, Tl, Si, Ge, Sn, Pb, As, Sb, Bi, Se, Te) or an element from Group IIB (Zn, Cd, or Hg). Note that these compounds could include the elements CNOPS as well. Compounds containing other metals were also considered, but there are very few such stable volatile compounds; while included for completeness, their stability to hydrolysis has not been systematically tested. The Extended CNOPSH List totals about 1500 and after a stability check is reduced to 862. By stable we mean the intrinsic stability of chemical bonds in molecules as well as hydrolytic stability (reactivity with water). Stability is further addressed in Section 2.4 and Section 4.1.1.

2.2.2. Database trawling for both the CNOPSH List and the Extended CNOPSH List. The second approach in con-

structing the list of CNOPSH molecules and the extended list of molecules (and also the list of halogenated compounds) that are stable and volatile was to trawl several different chemical databases (see Table A2) of physical properties of molecules (including silicon, germanium, and organometallic compounds) for any compound with a boiling point below 150°C (measured). The boiling point is included for our proxy for volatility. A cutoff temperature choice is made to limit numbers; the choice of 150°C is conservatively inclusive, as at STP any compound with a boiling point above 150°C will have very low vapor pressure and hence will likely be nonvolatile (Boublík *et al.*, 1973). A relatively small subset of the database trawling approach overlaps with the combinatorics approach (described above in Section 2.1.1), but most molecules were new to the list. In other words, these volatile molecules could include atoms other than CNOPSH, including silicon (and halogens, which are included in the list described in Section 2.3).

We also compiled a list of compounds provided by chemical suppliers (see Table A2) of molecular weight 150 or less. These compounds by definition are stable but are not necessarily volatile.

## 2.3. The Halogenated Compounds List

Halogenated compounds are treated separately for expediency in terms of search efficiency, because of their extensive numbers. Halogenated compounds fall into both organic and inorganic compounds. Inorganic compounds containing halogens were collected as part of the process for the Extended CNOPSH List, described in Section 2.2. Note, however, that the majority of inorganic halogenated compounds are not considered as stable for our purposes, as they are very reactive, particularly to water, and so are implausible atmospheric components. (There are a few exceptions, notably fluorinated compounds discussed below.)

The organic compounds containing halogens (halocarbons) are extensive in number. We constructed lists of members of 10 classes of organic chemicals and then exhaustively substituted the C-H bonds with C-halogen bonds (using Cl, Br, I, and F). For up to  $N=6$  non-H atoms, this resulted in a total of 10,463 compounds. Out of the total halogenated compounds, 3876 contained F.

## 2.4. Assessment for stability and volatility

Only stable molecules can accumulate in a planetary atmosphere and are potential biosignature gases. By stability we mean compounds that are stable on the order of days as pure entities at Earth's surface temperatures and pressures (STP) and are likely to be stable to reaction with water. To assess for stability, we take two approaches. First we consider stability of molecules by analogous groups (such as amines, esters, acid chlorides, etc., which are defined by their chemical reactivity), so we do not have to empirically determine the functional stability of every molecule. (Thus, for example, trimethylamine, ethyldimethylamine, methyl-diethylamine, etc. can be treated as one group with respect to stability.)

Next we assess stability of analogous groups of molecules at two levels: the stability against the reaction with water and the intrinsic stability of a pure compound. Reactivity

with water is critical because most biochemical reactions take place in water (*e.g.*, in the cytoplasm of the cell) and the immediate environment may also be water-based. Reactivity with water is determined by a literature search. Intrinsic stability of a class of molecules is evaluated by checking if its physical properties have been measured (*e.g.*, a particular melting or boiling point or IR spectrum experimentally measured). [Note that stability and volatility in hypothetical exoplanetary atmospheres is an application, not a property of the list, and must be assessed on a case-by-case basis and is an extensive endeavor (Section 4.2.1).]

There is no reliable way to predict volatility from theory. By volatility we mean that the partial pressure over a pure compound at STP is a substantial fraction of 1 atm. Boiling point is a convenient single estimator for volatility. We use measured boiling points when available. Accurate prediction of volatility and boiling points from theory is not possible, so where measured boiling points are not available we use the estimated boiling point as provided by the chemical software EpiSuite<sup>2</sup> (Stein and Brown, 1994).

As previously mentioned (Section 2.2.2), a cutoff temperature choice is made to limit numbers; the choice of 150°C is conservatively inclusive as at STP any compound with a boiling point above 150°C will have very low vapor pressure and hence will likely be nonvolatile. We also chose 150°C because of its relationship to stability; if a molecule is unstable, it will also not exist to be volatile. This is the temperature at which a wide variety of organic molecules become unstable to hydrolysis (Bains *et al.*, 2015) and hence the temperature at which life based on carbon chemistry in water becomes implausible (Kashefi and Lovley, 2003; Cowan, 2004). At temperatures significantly above 150°C, organic matter is degraded wholesale on a timescale of hours or minutes (Katritzky *et al.*, 1996). (Note that non-carbon-based life is beyond the scope of this paper.)

### 2.5. Assessment for production by life on Earth

To assess what fraction of the list of small, stable, and volatile molecules are produced by life on Earth, we took two separate approaches: a database approach and a manual literature search. The first approach was to automatically query two databases of molecules produced by life, ZINC<sup>3</sup> (Irwin and Shoichet, 2005) and the Dictionary of Natural Products<sup>4</sup> (Buckingham, 1993). These databases include references for the organism name from which a given molecule was isolated. We did not use the extensive UNPD<sup>5</sup> database (Gu *et al.*, 2013), because it does not easily include a method to go from molecule of interest to biological source information.

The manual literature approach involved review of summary papers, chemical classes produced by life, and a targeted search of the literature for specific chemicals, especially inorganic volatiles (because they are rarely incorporated into the databases). References used are listed in the tables, with examples provided in Tables B1, B2, and B3.

## 3. Results

### 3.1. A list of stable, volatile molecules up to $N=6$ non-H atoms

The main result of this work is the list of molecules that are stable and volatile at STP up to  $N=6$  non-H atoms. The list is divided into our three classes of CNOPSH molecules, an extended list of CNOPSH molecules, and a list of halogenated compounds. Sample lists are provided in the Appendix in Table B1, Table B2, and Table B3, respectively. Full lists are available at the URL noted in Section 5.

The tables contain the following information. The IUPAC name is included as a standard and unique reference name. The chemical structure information is provided in SMILES string format. The molecular weight is a convenient description of a molecule for additional filters or searches. The boiling point is our proxy for volatility. The boiling point, a relatively simple physical property, is not known for many molecules (for some molecules it has been predicted, indicated by a p=predicted vs. an e=experimental). A column to indicate whether or not the molecule is produced by life is followed by the reference for production by life.

The tables, as expected, contain a large number of organic molecules, a relatively small number of P-block element-containing molecules, very few volatiles that contain metals, and a large number of halogenated compounds. The sample tables provide illustrative entries of the diversity of compounds that could form volatile molecules at STP.

The total number of atoms in a molecule with  $N \leq 6$  non-H atoms varies greatly and depends on to what extent the molecule is reduced. Thus our total list of all categories contains molecules containing six non-H atoms that have as many as 19 (*e.g.*, methyl diethylamine:  $C_5H_{13}N$ ) or 20 atoms (hexane:  $C_6H_{14}$ ) and as few as 6 (tetrachloroethene:  $C_2Cl_4$ ).

Our list of volatile molecules has value in its applications in astrobiology. We now turn to one such application.

### 3.2. The fraction of molecules known to be produced by life on Earth

The list of all volatile, stable molecules is used to assess the fraction of molecules in the list that are known to be produced by life on Earth. This is partly to inform the search for biosignature gases and partly to further chemoinformatics. The tables (B1, B2, B3) indicate whether or not the molecule is known to be produced by life on Earth.

The value of the tables is in future applications (see discussion in Section 4.3), and here we provide a summary of our statistics.

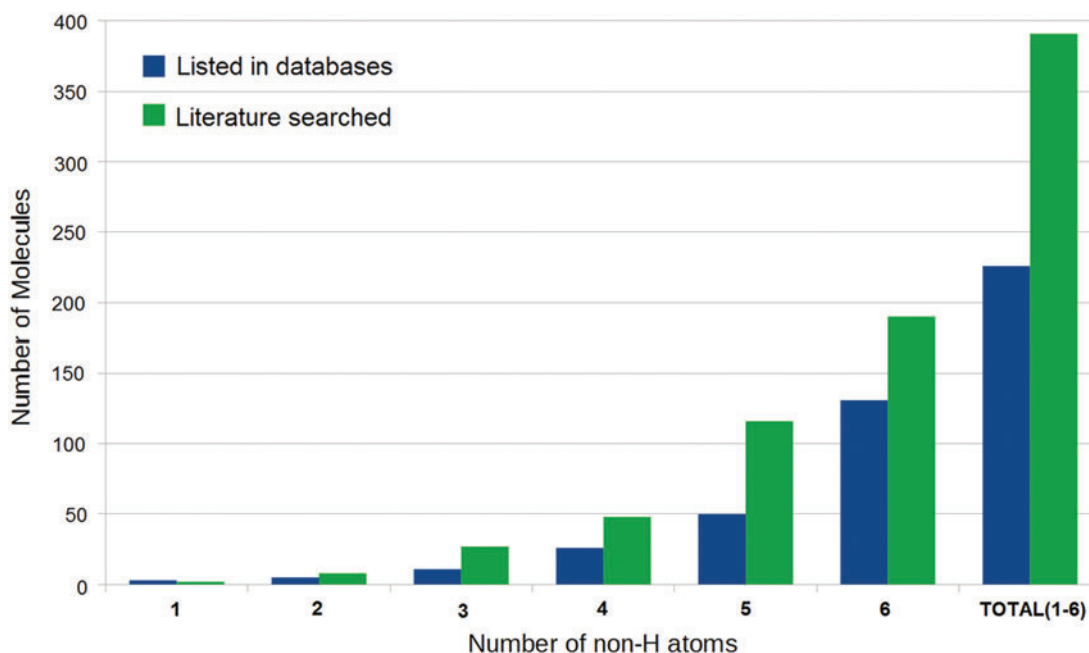
**3.2.1. CNOPSH life-producing molecules.** We surveyed all possible potentially volatile stable molecules of the six biogenic elements CNOPSH of up to  $N=6$  non-H atoms. We find that about 25% are known to be made by life. It is notable that about 63% of those produced by life were not listed in natural product databases (ZINC and DNP) but only recovered from a manual literature search (see Fig. 2), which may be because the existing databases are primarily oriented to pharmaceutical discovery rather than exhaustive cataloging metabolites. The molecules with non-H atoms  $N \leq 6$  are too small to provide enough specific points

<sup>2</sup><http://www.epa.gov/oppt/exposure/pubs/episuite.htm>

<sup>3</sup><http://zinc.docking.org>

<sup>4</sup><http://dnp.chemnetbase.com/intro>

<sup>5</sup><http://pkuxxj.pku.edu.cn/UNPD>



**FIG. 2.** Comparison of molecules produced by life as found in published databases and found by a manual literature search. For CNOPSH compounds only.

of interaction with pharmaceutical targets and are therefore not considered to be good candidates for development as drugs.

There may be yet more molecules produced by life that are not yet known. Experimental searches may find them; see Section 4.3.1.

**3.2.2. Extended CNOPSH life-producing molecules.** Life uses elements other than CNOPSH elements in a very limited number of compounds (*e.g.*, Berg *et al.*, 2002; Wackett *et al.*, 2004). For example, selenium is used widely by life but only for one metabolic function—as a component of glutathione peroxidase, an enzyme widely involved in cellular defense against toxins, metals, and reactive oxygen species (Burke, 1983). Another example is detoxification of heavy metals, which is carried out by their volatilization (*e.g.*, dimethyl mercury). It is not known why some P-block element-containing compounds are used and others are not (Thayer, 2002; Wackett *et al.*, 2004). For example, we found that life produces methyl germanes (Hirner *et al.*, 1998; Rosenberg, 2008) but not methyl silanes (Tacke, 1999), even though silicon and germanium have very similar chemistry (Greenwood and Earnshaw, 1997). We have found that 43 compounds falling into our Extended CNOPSH List are made by life, out of 862 total. (See Table B2.)

**3.2.3. Halogenated compound life-producing molecules.** Life produces a diverse set of compounds with Cl, Br, or I bonded to a carbon atom. These include small molecules such as methyl chloride (which has already been considered as a biosignature gas; Segura *et al.*, 2005) as well as much more complicated molecules such as the antibiotic tetracycline.

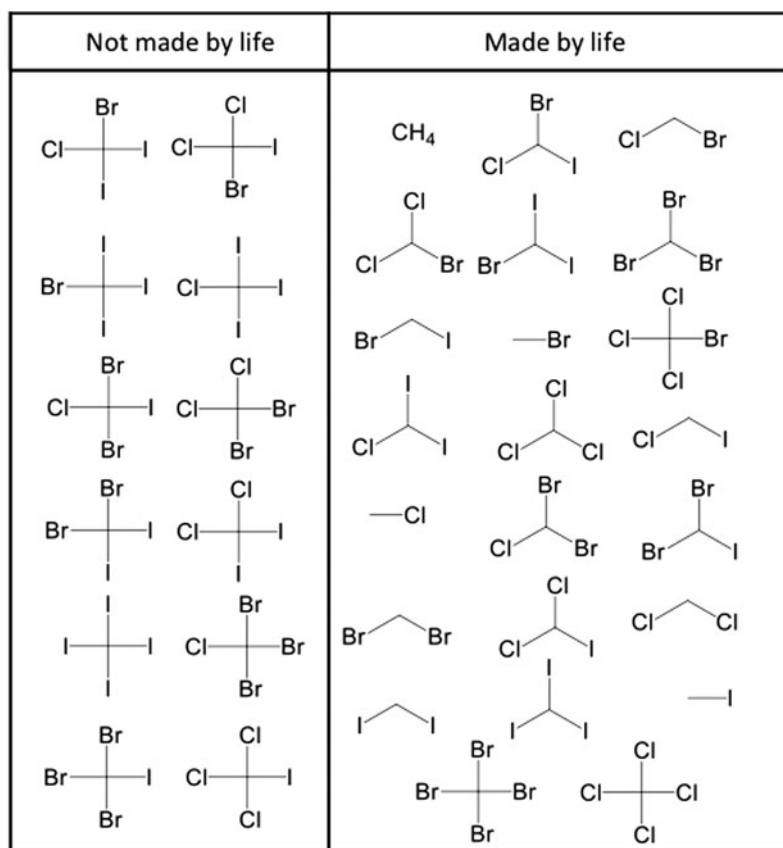
Life does not produce all small organohalogen molecules. There are three categories of halogenated compounds not

known to be produced by life. The first category is molecules that are structurally similar to molecules that are known to be produced by life but that have not been detected experimentally. This category is for compounds with Cl, Br, I bonded to carbon. For example, many of the methyl halides are known to be made by life, but some are not (Fig. 3 and see sample Table B3). There does not appear to be any physicochemical or chemical pattern to predict which are made by life and which are not (Gribble, 2003; Paul and Pohnert, 2011). We could speculate that a more exhaustive experimental search for these compounds could expand the list of halocarbons produced by life. Many of them can be easily overlooked, especially those that are produced in trace amounts or by rare sets of organisms.

The second category of halogenated volatiles not produced by life is fluorine compounds. Life produces very few organofluorine compounds, and their production appears to be very specialist biochemistry restricted to a small number of species [such as fluoroacetone (*e.g.*, Walker and Chang, 2014)].

The third category is the halogens bonded to any atom other than carbon. Such bonds are usually quite hydrolytically unstable (highly reactive with water, such as the P-Cl bond) or extremely reactive (such as oxides of chlorine or fluorine) so would not be expected to be found as stable chemicals accumulating in a planetary atmosphere. Hypochlorite (chlorine bonded to an oxygen) is one exception of a stable molecule in this category.

Of the 5708 CNOPS compounds that contain Cl, Br, or I, 103 are known to be made by life. Only 3 of 206 inorganic halogenated compounds are known to be made by life (Table B3). However, the halogenated compounds reported are those that are made in the largest amounts by easily accessible species, usually seaweeds (Gribble, 2003; Cabrita *et al.*, 2010; Paul and Pohnert, 2011; Pomin, 2011). Out of



**FIG. 3.** Halomethanes (excluding F) produced and not produced by life. A halomethane compound is a derivative of methane ( $\text{CH}_4$ ), where one or more of the H atoms are replaced by a halogen atom. A carbon atom is implicit in the structures shown in this figure. Out of the 34 (non-F) halomethanes, 12 are not known to be produced by life. Why just 22 are produced by life is not known.

the total 4109 F-containing compounds<sup>6</sup>, only three are made by life [fluoroacetate (Moss *et al.*, 2000; Murphy *et al.*, 2003), fluoroacetone (Peters and Shorthouse, 1967), fluoroacetaldehyde (Moss *et al.*, 2000)]. Given that over half of all possible chloro-, bromo- and iodomethanes are made by life (Fig. 3), it is possible that many more halogenated biochemicals await discovery.

#### 4. Discussion

##### 4.1. Caveats and future extensions to the list

4.1.1. Challenges for assessing stability and volatility. For this work we have considered an approximation that all small molecules can be considered to be volatile. Here we review the caveats to our approach, first starting with some background.

A compound is volatile if the interactions between the molecules are of similar energy to the thermal energy of the molecules. Volatility is also influenced by noncovalent interactions. Noncovalent interactions between molecules are van der Waals-type forces, dipole interactions including hydrogen bonds, and electrostatic or charge interactions. Van der Waals interactions to a first order are dependent on molecular weight, so small molecules tend to be volatile. Dipole interactions are dependent on structure, but for small molecules even the most highly polarized molecules such as HF

are still volatile. Charges on molecules render them nonvolatile (at least at ambient temperatures). However, for most organic molecules that have a charge at neutral pH, there is a pH at which they are uncharged and hence volatile. (This is not true of quaternary ammonium salts and zwitterions.)

Volatility was assessed by taking measured or calculated boiling points at standard pressure from ChemSpider. For searching ChemSpider and for generating combinatorial sets of halogens, we used a cutoff of a boiling point of 150°C to limit the number of compounds considered and for stability reasons (Section 2.4). The adopted cutoff is a conservative cutoff, such that further work with larger molecules than those included in our list of all small molecules should be investigated more carefully. For smaller data sets, including literature searches, we truncated searches based on molecular size and manually checked that the compounds were volatile based on boiling point.

The 150°C cutoff for volatility is limiting because we may have included molecules that may have low volatility at STP. The cutoff could be replaced by a molecule-by-molecule effort to evaluate its vapor pressure at STP. Individual molecules of choice will have to be evaluated in detail for the specific environment of interest, especially for non-STP conditions for exoplanetary environments different from Earth.

Turning to stability, we used a manual literature search for inherent stability and stability to hydrolysis. Stability is not easily predictable from chemical structure, but this may be less of a problem for larger organic molecules than for the small ones we worked with, as one can argue the stability of a large molecule by analogy with smaller analog molecules.

<sup>6</sup>More than the halogen combinatorics number because there are some F compounds in the chemical suppliers databases as well.



Stability to hydrolysis depends on the rate at which the most labile bond in a molecule is cleaved by water. For many molecules it is obvious from inspection that there is no water-labile bond present (easily attacked by water). For example, the halocarbons are deemed to be inherently stable to hydrolysis, as the  $sp^3$  carbon-halogen bond is inherently stable to hydrolysis. For others, it is clear that there is a bond that will be rapidly attacked by water, for example, the silicon-chlorides. Stability and instability of some classes of molecules can therefore be determined by inspection. For other molecules, stability to hydrolysis cannot be easily determined by inspection. For these we conducted a literature search to look for experimental measures of stability or instability. The challenge is how to assess the class of molecules intermediate between stable and unstable, particularly in the absence of data, and how to extend our understanding of stability to stable or unstable molecules at conditions other than STP. For example, at the bottom of the atmosphere, conditions on Earth approximate STP, and hydrolysis is dominant. However, at the top of the atmosphere, pressure and temperature differ substantially from the surface, and photolysis is the dominant destruction mechanism for most molecules.

We chose STP because this is where most properties are measured or calculated. While STP is appropriate for terrestrial-based applications of our list of molecules, this is a limitation for exoplanets whose environments will differ from STP. Stability and volatility away from STP will have to be estimated or measured for specific biosignature gases of interest; fortunately the range of temperatures and pressures is not infinite but largely constrained to those that support liquid water. Extensions beyond STP are a huge and demanding piece of research that we hope will be initiated in the future.

We consider stability to hydrolysis to be an important feature for a biosignature gas for mainstream astrobiological studies. The consensus is that water is the most likely solvent for life and so will be present on the surface of any inhabited world. In more detail, if we assume that life is based on water, then any volatile molecule will be made in water and will have to diffuse out of the cell in which it is made. While the instability of a molecule in life's surface surroundings and air is also a factor, it is likely not the main factor. Life's existence in more exotic solvents than water has been suggested, with the solvents including water/ammonia mixtures (Fortes, 2000), liquid methane or nitrogen (Bains, 2004; McKay and Smith, 2005), and supercritical  $CO_2$  (Budisa and Schulze-Makuch, 2014). If these alternative solvents are to be considered seriously, then reactivity to the appropriate solvent will have to be substituted for reactivity to water. This is work for the future.

**4.1.2. Completeness of the list.** In construction of a database, one must ask if anything is missed, either important categories or a substantial number of entries within an established category. In terms of search method, our multi-pronged approach should not have missed any molecules in our computational, database, and literature sources (see Fig. 1). Out of the 85,000 or so references collected, we found about 68,000 unique structures, out of which 14,332 were  $N \leq 6$  non-H atoms.

Any missing molecules may be from the challenge of assessing volatility and stability. Here we aimed on the con-

servative side by including molecules that may have limited volatility or stability.

List completeness is not considered a problem for our artificial cutoff for molecules with  $N \leq 6$  non-H atoms. It may well be a big challenge for a complete list of all volatile and stable molecules at STP, when applications in chemical space require completeness and where the manual multi-pronged search approach may not be practical.

**4.1.3. Extending the list.** The validation process of identifying chemicals and searching for stability measures, physical processes, and production by life is very labor intensive. While some processes may be automated (generating molecular structures is straightforward via computation), others are more challenging. Physical properties that are not known cannot be easily calculated. Finding out from the literature whether or not a chemical is known to be made by life is difficult because the literature is not indexed to be searched by chemical structure and the databases of molecules made by life are quite incomplete (Fig. 2). Even identifying the papers requires expertise and is not just a matter of search terms. Understanding which chemicals in the papers are genuinely biological products also requires substantial expertise. Distinguishing between a true natural product (chemical that is a product of a normal metabolism of a living organism) and a potential metabolite of an industrially synthesized chemical such as a drug or a pesticide can sometimes be challenging (see the example of aminourea in Van Poucke *et al.*, 2011)<sup>7</sup>. We advocate for a community effort to register biological molecules.

**The CNOPSH List.** The number of compounds increases nearly exponentially (based on combinatorics) with the number of non-H atoms. This is because replacing a hydrogen with one new atom from CNOPS increases the opportunity to add more atoms; that is, the rate of change of the number of atoms is a function of the number of atoms. A completely exhaustive combinatorics method is one method, with a series of rules and checks to screen for real molecules. This approach is impractical for our purposes, as it generates trillions of molecules. A more practical approach might be combinatorics based on substructures that can be combined, or using genetic algorithms of graph theory. There are such established computational techniques for drug design for CNOSH molecules, such as those used for the Chemical Universe Database (Reymond *et al.*, 2012), which results in nearly 1 billion structures. Generally the larger/heavier the molecule is, the less likely it is to be volatile. Although boiling point measurements usually are not available and rigorous calculation is very difficult, estimates are straightforward.

**The Extended CNOPSH List.** Our approach to inorganic compounds and others in the Extended CNOPSH List so far was manual and could be extended by a combination of

<sup>7</sup>Van Poucke *et al.* (2011) proved that aminourea (semicarbazide) is a natural product produced by prawns that accumulates predominantly in their exoskeleton. It was unclear if aminourea is a natural product as it is also a metabolite of a pesticide nitrofurazone. Prawns exposed to nitrofurazone accumulated aminourea in their exoskeletons in much higher amounts than those not exposed, which produce aminourea naturally. This was later confirmed by other groups as well.

manual and computational methods. Our existing list could be extended by replacing H atoms with organic groups (or halogens). As a complement we could take our current list of organic compounds, take every C-H bond, and replace the H with any P-block element. In any case, we would expect to reach the volatility limit relatively quickly, but volatility would be difficult to calculate.

*The Halogenated Compounds List.* The Halogenated Compounds List would be extended from the CNOPS List as already described (Section 2.3). Prediction of the boiling points is fairly reliable for hydrocarbons, the largest class. We must emphasize the anticipated staggeringly large number of fluorinated compounds that will be stable and volatile. There are (for example) five hexanes (C<sub>6</sub>H<sub>14</sub> compounds: *n*-hexane, 2-methyl pentane, 3-methyl pentane, 2,2-dimethylbutane, and 2,3-dimethylbutane) but 2460 different hexane fluorocarbons, all of which are expected to be stable to heat and hydrolysis and to have a boiling point <80°C.

*Production by life.* A manual literature search for production by life is the only option at this point (Section 2.5). This is the limiting factor for making progress in any application described below.

#### 4.2. Potential applications for astrobiology

4.2.1. Exoplanets: path forward for a list of potential biosignature gases. We expand on our initial motivation for constructing the list of all volatile, stable molecules: the concept that any gas could potentially be a biosignature gas accumulating in the atmosphere of another world. This concept is supported by many different examples.

Chemicals produced as a result of secondary metabolism (metabolism not related to acquiring energy or biomass buildup) can have very diverse functions that depend on an evolutionary history of a species. More specifically, volatile chemicals produced by life as a result of secondary metabolism could be used for signaling many behavioral cues like aggression, defense, sexual attraction, trail following, and other means for interaction with the environment [thousands of examples are given in the Pherobase<sup>8</sup> (El-Sayed, 2014)]. Many volatile molecules produced as carriers of such signals are unique and originated from coevolution among species; hence it is impossible to predict which specific volatile molecule will be the carrier of which behavioral signal (*e.g.*, Tirindelli *et al.*, 2009; Deisig *et al.*, 2014; Steiger and Stökl, 2014). An illustration of such a multispecies, intricate, and complex volatile chemical signaling network was discovered recently between apple trees, great tits (*Parus major*), and caterpillars of *Operopthera brumata* (Amo *et al.*, 2013). Great tits respond to specific volatile molecules (dodecanal and alpha-farnesene) emitted by apple trees infested by *Operopthera brumata* (Amo *et al.*, 2013). Attracted to the trees, the great tits eat the caterpillars. The identity of volatile chemicals produced by caterpillar-infested trees and those that are free of them is very different and specifically tailored to *Parus major* olfactory receptors (Amo *et al.*, 2013).

Methyl chloride (CH<sub>3</sub>Cl) and dimethyl sulfide (CH<sub>3</sub>SCH<sub>3</sub> or DMS) have been studied as exoplanetary biosignature gases (Segura *et al.*, 2003, 2005; Scalo *et al.*, 2007; Domagal-

Goldman *et al.*, 2011; Rugheimer *et al.*, 2013; Seager *et al.*, 2013a, 2013b; Rugheimer *et al.*, 2015) and are additional examples of the diversity of gases produced by life whose production is not linked in any predictable way to the physical or chemical properties of Earth. Rather, they happen to be produced in relatively large amounts by organisms that are also relatively common (Bates *et al.*, 1993; Kiene *et al.*, 2000; Yoshida *et al.*, 2004).

Dimethyl sulfide is believed to be produced overwhelmingly by oceanic plankton from dimethylsulfoniopropionate (DMSP; Alcolombri *et al.*, 2015), but other routes such as the methylation of H<sub>2</sub>S or methanethiol by photosynthetic, partially anaerobic bacterial mats (Visscher *et al.*, 2003) are minor sources on Earth and may be more substantial routes of production on other worlds. DMSP itself is mainly made by oceanic phytoplankton, although the reason they make DMSP is controversial (reviewed in Sunda *et al.*, 2002; Brimblecombe, 2003), but DMSP is also made by macroalgae (Stefels, 2000) and some species of land plants (Otte *et al.*, 2004). Some of those land plants are associated with bacteria that produce DMS from the DMSP (Todd *et al.*, 2007; Johnston *et al.*, 2008), but it is unclear why the plants make DMSP (reviewed in Otte *et al.*, 2004), and the reason is probably different from the reason that marine phytoplankton make it. This emphasizes that the biochemistry and ecology of a potential biosignature vary even among Earth life and so cannot be assumed to be the same on other worlds.

Methyl chloride is specifically synthesized by an enzyme system that halogenates small hydrocarbons (Ni and Hager, 1998; Vaillancourt *et al.*, 2006). Although the biosynthetic pathways are well understood (Rhew *et al.*, 2003; Roje, 2006), the exact ecological role and physiological function of methyl chloride remain to be determined, despite earlier suggestions that methyl chloride is produced as a toxicant to deter predators and/or suppress competitors (Hartmans *et al.*, 1986; Manley, 2002). Methyl chloride is widely produced by many plant and algal species (Wuosmaa and Hager, 1990; Nadalig *et al.*, 2011; Rhew *et al.*, 2014) and also by fungi and marine bacteria (Tait and Moore, 1995; Khalil and Rasmussen, 1999).

While some biogenic gases and biochemicals can be predicted from the thermodynamics of metabolism, it is not straightforward to predict more complex biochemicals, such as hormones and others used in ecological dynamics. Gases that are oddities of biochemistry (*i.e.*, gases that are members of groups of chemicals with very few members made by life and that are usually made by few species) further support the concept that a variety of gases might be produced by life even in a planetary environment with known composition. For example, fluoroacetone is one of the few F-containing compounds made by life, and it is made by only a few species (see the reviews in O'Hagan and Harper, 1999; Gribble, 2002; also discussed in Section 3.2.3).

We now turn back to the path forward for exoplanets. To proceed, we envision taking molecules, or classes of molecules, and assessing (1) how likely they are to accumulate in an exoplanetary atmosphere, (2) whether or not they are spectroscopically detectable by a remote space telescope, and (3) whether or not there are false positives for the gas in question, in context with different exoplanetary atmosphere

<sup>8</sup><http://www.pherobase.com>

and surface conditions, and whether any other spectral features could be observed to support or refute a given gas as being of biogenic origin. The first point depends on the UV radiation environment from the host star that drives photochemistry, the atmospheric composition and mass, and the surface and ocean chemistry (including sources and sinks). A practical path forward is to take classes of molecules and determine if they are stable and volatile at non-STP conditions appropriate to the types of exoplanets under consideration and can accumulate in different planetary environments by integration into existing models of planetary chemistry and photochemistry.

The spectroscopic detectability of gases relies on molecular lines or bandhead estimates, which for many molecules do not exist yet. For example, the renowned HITRAN compilation of molecular line lists and cross sections (Rothman *et al.*, 1998) includes about 50 molecules, those relevant for studies of Earth's atmosphere. A useful collection for exoplanets from the Virtual Planet Laboratory<sup>9</sup> has about 130 molecules [compiled from HITRAN, PNNL<sup>10</sup> (Sharpe *et al.*, 2004), and personal collections]. A few expert research groups calculate molecular line lists from *ab initio* theoretical quantum mechanics calculations (*e.g.*, Tennyson and Yurchenko, 2012), efforts that can take a year or more per molecule. In preliminary work (Zhan *et al.*, unpublished data), we have found gas phase spectra for about 1000 of the 14,000 volatiles in our list. Most of our spectra come from the IR gas phase NIST online infrared database<sup>11</sup> (Sharpe *et al.*, 2004), which is widely distributed and is the basis for most of a few dozen or so commercial online spectral databases. About one-third of the 1000 spectra on our list of volatile molecules found in NIST are actually digitized transmittance spectra data taken experimentally from decades ago (The Coblentz Spectral Collections<sup>12</sup>) and are lacking path length or other information, preventing cross sections from being derived, and therefore are not useful as input to exoplanet model atmosphere codes. Many of the spectra noted in the above databases are limited in wavelength range. More efforts to calculate even crude spectral information are needed to advance the search for biosignature gases. Our database of all molecules will eventually include links to existing spectra.

Clearly, there is a long way to go in coming up with a list of all potentially useful biosignature gases. The concept for application to potential biosignature gases is illustrated in Fig. 4.

The first chance for the search for biosignature gases on exoplanets is with the James Webb Space Telescope (JWST), scheduled for launch in 2018 (Gardner *et al.*, 2006). JWST will be able to observe the atmospheres of small transiting exoplanets transiting small stars (*e.g.*, Deming *et al.*, 2009; Beichman *et al.*, 2014). A dozen or more habitable zone planets are anticipated to be discovered by ground- and space-based surveys including the upcoming MIT-led NASA mission TESS, scheduled for launch in 2017 (Ricker *et al.*, 2014). With its capability of near-IR spectral resolution of a few

hundred to a few thousand, JWST will be capable of observing life-produced gases, if they exist. But which gases are best to search for out of our list of molecules will require in-depth study (Zhan *et al.*, 2015).

**4.2.2. Nonterrestrial biochemistry.** Our list of all volatile, stable molecules is constructed based on the assumption that life uses water as a solvent but is otherwise agnostic to the chemistry of life. Specifically, life originating in very different physical or chemical environments might select different basic sets of atoms and bonds from which to build biochemistry. For example, life on Earth rarely uses the C-F bond (Gribble, 2002; Wackett *et al.*, 2004), which is one of the strongest single bonds known. Life evolving at much higher temperatures could use more C-F bonds to compensate for the greater instability of molecules to hydrolysis at those temperatures, and consequently would generate fluorocarbon biosignature gases. Fluorocarbons are particularly interesting as signature molecules as they are anomalously volatile for their molecular weight.

Small, covalent molecules are built from atoms with between 2 and 5 valencies, which can form the networks of bonds needed to form different structures. However, if they are not to form unbounded structures (polymers), small molecules need monovalent atoms to “cap” the valencies. On Earth, hydrogen nearly always plays this role. Life in extremely low water environments might choose chlorine as a monovalent atom in molecule construction rather than hydrogen, a possibility drawn from studies that address the point of life in low-water-activity environments (Grant, 2004; Schulze-Makuch and Irwin, 2008) and work describing a hypothetical chlorinic photosynthesis (Haas, 2010).

We do not propose the above as specific biochemistries for which to search. Rather, we illustrate that we can conceive of environments where very different chemistry might make sense. In order to be prepared for the biosignatures that we would need to detect to find evidence for life in such environments, we want to be inclusive in the chemical space explored.

#### 4.3. Potential applications to terrestrial biochemistry

Ultimately, we would like to be able to understand more precisely what a biochemistry on another world would look like, including the by-product gases detectable by a distant observer, beyond the speculations given above. Outcomes from this work suggest that such predictions might be at least partly feasible and that higher-order abstractions about the chemistry of life may be possible with an organized structure list of molecules produced and not produced by life.

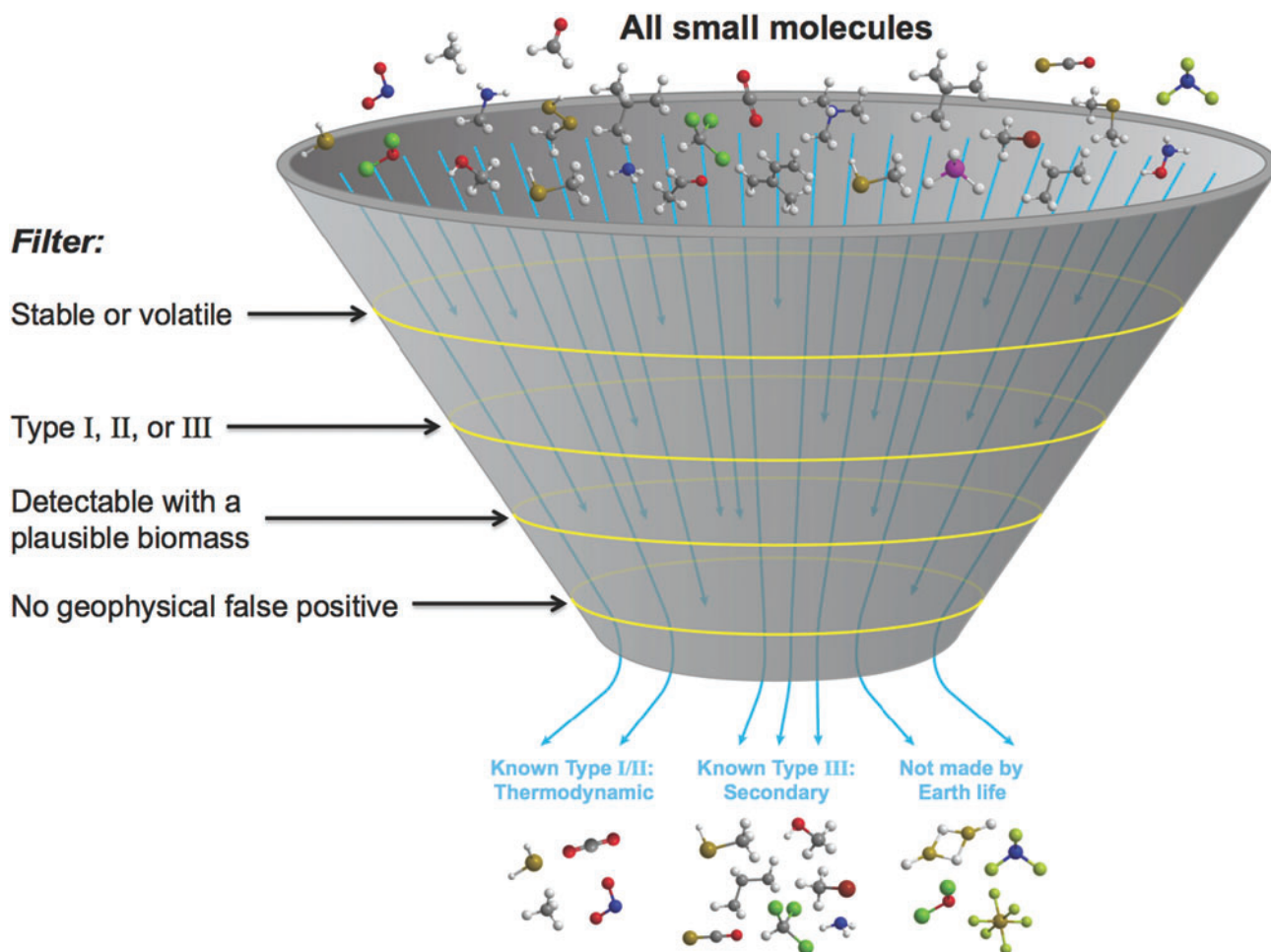
**4.3.1. Yet undiscovered metabolites.** One basic application is to ask why some metabolites are *not* made by Earth life. Our list shows some unexpected gaps in our knowledge of molecules produced by life, molecules that are “missing” from the database. By “missing” we mean molecules not produced by life among a set of many structurally similar molecules that are known to be made by life. For example, there is no report of terrestrial life making CCl<sub>3</sub>I, although life does make CCl<sub>4</sub>, CBr<sub>4</sub>, and CHCl<sub>2</sub>Br (see Fig. 3). There is only one report of life making dimethyl ether (CH<sub>3</sub>-O-CH<sub>3</sub>), although many organisms make other volatile ethers.

<sup>9</sup><http://vpl.astro.washington.edu/spectra>

<sup>10</sup>[nwir.pnl.gov](http://nwir.pnl.gov)

<sup>11</sup><http://www.nist.gov/srd/nist35.cfm>, <http://webbook.nist.gov/chemistry>

<sup>12</sup><http://www.coblentz.org/education/spectral-databases>



**FIG. 4.** Schematic for the concept of considering all volatile molecules in the search for biosignature gases. The goal is to start with chemistry and generate a list of all small molecules and filter them for utility as biosignature gases. The first filter is for molecules that are stable and volatile in temperature and pressure conditions relevant for exo-Earth planetary atmospheres. Further filters relate to the gas detectability, aided by its Type classification and spectroscopic characteristics. Geophysically or otherwise generated false positives must also be considered. In the ideal situation, this overall conceptual process would lead to a finite but comprehensive list of molecules that could be considered in the search for exoplanetary biosignature gases, based on atmospheric models with surface source and sink input details, as well as strengths of spectroscopic features of molecules. Figure credit: S. Seager and D. Beckner. Figure originally published in Seager and Bains (2015).

Focused research in the field or lab to search for organisms producing these molecules could be informative for the biochemical space of molecules produced by life.

4.3.2. Patterns in terrestrial biochemistry's use of elements. We notice that terrestrial life does not use all elements with equal frequency. We quantified the use of elements by life as follows. We postulate that the probability that a compound containing a particular combination of elements is made by life is the product of element-specific probabilities. Thus, for example, carbon and nitrogen have high probabilities (0.71 and 0.32, respectively), and fluorine has a very low probability (0.01). Purely based on these probabilities, the chances that a hydrocarbon or an amine is made by life are therefore high, a fluorocarbon low, and the chances that nitrogen trifluoride is made by life are essentially zero. Predictions of how many compounds containing a set of elements are made by life using these probabilities match the

observed frequencies to  $r^2=0.92$ . In general, CHONS are commonly used; other elements are used less commonly depending (roughly) on their "distance" on the periodic table from these core elements. We do not have a theoretical explanation for this, apart from speculating qualitatively that the chemistry of life is adapted to handle carbon, hydrogen, oxygen, nitrogen, and sulfur atoms; hence the more chemically different from CHONS an element is, the more adaptation of that basic chemistry is required to handle that element (Petkowski, Bains, Zhan, and Seager, unpublished data). Notably, phosphorus is not in this high-probability group; despite its presence (as phosphate) in many metabolites, the chemical diversity of phosphorus biochemicals is very limited (Petkowski, Bains, Zhan, and Seager, unpublished data). Our observation shows that a systematic compilation of all molecules can reveal patterns about the chemistry of life that are different from those made by calculating the bulk composition of life (Chopra and Lineweaver, 2010)

and might inform our understanding of why life has the chemistry that it does. Further work to explore these patterns is underway.

4.3.3. Patterns in chemical motifs for chemical scaffolding. The list of volatile, stable molecules as a type of systematic survey can also help identify patterns in Earth life's use of chemistry. In our analysis for the fraction of molecules in our list produced by life, we found that the large majority of the small molecules that life does not make are nevertheless made by living organisms as a chemical substructure contained in other, larger molecules. We call the chemical substructure, or fragment, a "chemical motif." A good example of a chemical motif is dithioformic acid ( $\text{CH}_2\text{S}_2$ )—unstable on its own but present in many biological molecules as a chemical motif (Lim *et al.*, 1998). We have found that including chemical motifs in the production of molecules by life more than doubles the fraction of molecules produced by life for the CNOPSH molecules of non-H atoms.

We have further found that molecules or motifs rarely or not at all produced by life appear to fall into distinct categories. For example, we have found that some chemical motifs are very rarely made by life, such as allenes and cumulenes (an allene is a molecule with two consecutive carbon atoms double bonded to each other; a cumulene is a molecule with three or more consecutive carbon-carbon double bonds). We have also found that some chemical motifs are never made by life, such as triply bonded phosphorus [P(III), trivalent phosphorus<sup>13</sup>]. We believe that further analysis of a complete list of motifs cross referenced to our list of all small molecules, via chemical motifs, can have use in toxicity and pharmacology studies (Petkowski, Bains, Zhan, and Seager, unpublished data).

## 5. Summary

We have constructed a list of molecules up to  $N=6$  non-hydrogen atoms that are stable (in the presence of water) and volatile at STP. The list contains about 14,000 molecules: about 2500 are composed of the six biogenic elements, CNOPSH; about 900 are inorganics; and about 11,000 are halogenated compounds (a large set because of the large number of combinatorial possibilities of adding halogens to carbon skeletons). The list was constructed by a combinatorial approach and an intense database and literature search.

We further investigated which of the molecules in our list are known to be produced by life on Earth. About one-quarter of the compounds containing the six biogenic elements up to  $N=6$  non-hydrogen atoms are known to be produced by life on Earth. Very few of the inorganics are produced by life. A very small fraction of halogenated compounds are known to be produced by life. Even though dozens are found to be produced, for example, by seaweed species, there are thousands of possible  $N=6$  non-H-atom halogenated organic volatile molecules.

More specifically, for molecules composed of the six biogenic elements for  $N \leq 6$  non-hydrogen atoms produced by

life on Earth, we found that about one-quarter (specifically, 622) were produced by life on Earth. As an aside, via this search we found that database sources of chemicals made by life are substantially incomplete for small molecules (Fig. 2). Our manual literature search revealed that about 60% of the biogenic compounds were not in existing online databases. This reveals a need in the scientific community for a better registry system for biogenic molecules.

The applications of the list are twofold. One is related to the future search for biosignature gases in exoplanetary atmospheres, and that is to identify all molecules that can accumulate in different types of hypothesized exoplanetary atmospheres that also have strong spectral signatures, with future work on false positives in an exoplanetary environmental context. The motivation for the biosignature gas application is that many gases on Earth that have been studied in the context of exoplanets (such as methyl chloride) appear to depend on accidents of ecology and the whims of evolution (such as fluoroacetone or stibine  $\text{SbH}_3$ ) and so may be very different on other worlds. Hence we are motivated to construct a list of all molecules that are stable and volatile to be considered as potential biosignature gases on exo-Earths.

Related to the search for biosignature gases, we stated that Earth's atmospheric gases to the parts-per-trillion level by volume are all produced by life (with the exception of the noble gases and the set of fluorocarbons that are entirely anthropogenically produced), though their dominant source may be abiotic.

The second application is for the search for patterns among the molecules or motifs produced or not produced by life for understanding of limits of terrestrial biochemistry. Given that only a few hundred organisms have been studied for their production of volatiles, out of the (probable) millions of species on the planet, this suggests that many more chemicals remain to be discovered as gases produced by life on Earth.

Our database of molecules is available for community use. Our current iteration of the database is in a form that allows for detailed classification of molecules with respect to included basic physicochemical properties and/or production by life. Yet the list of molecules is incomplete, both because of our cutoff factors in construction (including our proxy for volatility, STP, and the  $N=6$  non-H-atom cutoff) and the lack of information on biological occurrence. To this regard, the main future work is extending the database to larger molecules, for an exhaustive list of stable and volatile molecules. We emphasize the need for a community effort to register all discovered biological molecules in a single database. The community can download and also suggest additions to our database provided at <http://www.allmolecules.org>.

## Acknowledgments

We thank MIT and the MIT Amar G. Bose Research Grant for support. We thank the reviewers for comments that improved the manuscript. We thank Charles Darrow, Victor Pankratius for useful comments, Ehsan Tofigh for computational support in the early phase of this work, and Dr. Jozica Dolenc from ETH Zürich, Informationszentrum (<http://infozentrum.ethz.ch>) for help with boiling point calculations.

<sup>13</sup>Claims that phosphine is produced by anaerobic organisms directly, rather than indirectly by biologically produced acid attack on minerals or human-made iron metal, remain controversial.

## Appendix A

TABLE A1. EARTH'S ATMOSPHERIC COMPONENTS

<i>Compound name</i>	<i>Formula</i>	<i>Typical atmospheric concentration</i>	<i>Primary terrestrial atmospheric source</i>	<i>Example of biological production</i>	<i>Reference for biological production</i>
Nitrogen	N <sub>2</sub>	78%	—	Denitrifying bacteria	1
Oxygen	O <sub>2</sub>	21%	Photosynthesis	Photosynthesis	1
Water	H <sub>2</sub> O	1–4%	Evaporation	Respiration	1
Argon	Ar	9340 ppm	Outgassing	<i>Not produced</i>	—
Carbon dioxide	CO <sub>2</sub>	350 ppm	Outgassing, biology, anthropogenic	Respiration	1
Neon	Ne	18.18 ppm	Outgassing	<i>Not produced</i>	—
Helium	<sup>4</sup> He	5.24 ppm	Outgassing	<i>Not produced</i>	—
Methane	CH <sub>4</sub>	1.7 ppm	Biology	Methanogenesis	1
Krypton	Kr	1.14 ppm	Outgassing	<i>Not produced</i>	—
Hydrogen	H <sub>2</sub>	0.55 ppm	H <sub>2</sub> O photolysis	Hydrogenase H <sub>2</sub> production in phototrophs	2
Nitrous oxide	N <sub>2</sub> O	320 ppb	Biology	Ammonia oxidation (nitrification)	3
Carbon monoxide	CO	125 ppb	Photochemistry	Mammalian CO signaling	4
Xenon	Xe	87 ppb	Outgassing	<i>Not produced</i>	—
Ozone	O <sub>3</sub>	10–100 ppb	Photochemistry	Inflammation in animals	5
Hydrogen chloride	HCl	~1 ppb	Sea salt	Halocarbon metabolism (as chloride)	6
Isoprene	C <sub>5</sub> H <sub>8</sub>	1–3 ppb	Plants	Trees	7
Hydrocarbons, <i>e.g.</i> , ethane	C <sub>2</sub> H <sub>6</sub>	0.2–3 ppb	Fires, oceans, anthropogenic	Oceanic bacterial metabolism	8
Benzene and other aromatics	C <sub>6</sub> H <sub>6</sub>	0.1–1 ppb	Anthropogenic	Made by mushrooms, trees	9
Ammonia	NH <sub>3</sub>	0.1–3 ppb	Biology	Nitrogen fixation by many bacteria	10
Nitric acid	HNO <sub>3</sub>	0.04–4 ppb	Photochemistry	Nitrifying bacteria (as nitrate)	11
Methyl chloride	CH <sub>3</sub> Cl	612 ppt	Biology	Oceanic bacteria	12
Carbonyl sulfide	OCS	500 ppt	Biology	Lichens in soils	13
Nitric oxides	NO, NO <sub>2</sub>	30–300 ppt	Biology	Ammonia-oxidizing bacteria	11
Difluorodichloromethane	CCl <sub>2</sub> F <sub>2</sub>	300 ppt	Entirely anthropogenic	Anthropogenic	1
Trichlorofluoromethane	CClF <sub>3</sub>	178 ppt	Entirely anthropogenic	Anthropogenic	1
Trichloroethane	CH <sub>3</sub> CCl <sub>3</sub>	157 ppt	Anthropogenic	Seaweed	14
Tetrachloromethane	CCl <sub>4</sub>	121 ppt	Anthropogenic	Seaweed	15
Tetrafluoromethane	CF <sub>4</sub>	69 ppt	Entirely anthropogenic	Anthropogenic	1
Chlorodifluoromethane	CHClF <sub>2</sub>	59 ppt	Entirely anthropogenic	Anthropogenic	1
Hydrogen sulfide	H <sub>2</sub> S	30–100 ppt	Biology	Sulfide-reducing bacteria	16
F113	C <sub>2</sub> Cl <sub>3</sub> F <sub>3</sub>	30–40 ppt	Entirely anthropogenic	Anthropogenic	1
Dichloromethane	CH <sub>2</sub> Cl <sub>2</sub>	30 ppt	Anthropogenic	Mushrooms	9
Dichloroethane	C <sub>2</sub> H <sub>4</sub> Cl <sub>2</sub>	26 ppt	Anthropogenic	Marine bacteria	17
Bromomethane	CH <sub>3</sub> Br	22 ppt	Biology	Seaweed	12
Sulfur dioxide	SO <sub>2</sub>	20–90 ppt	Volcanic, anthropogenic	Sulfur disproportionating bacteria (as sulfite)	18
Chloroform	CHCl <sub>3</sub>	16 ppt	Anthropogenic	Marine bacteria	12
Carbon disulfide	CS <sub>2</sub>	15 ppt	Anthropogenic	Tree roots and leaves	19
F114	C <sub>2</sub> Cl <sub>2</sub> F <sub>2</sub>	14 ppt	Entirely anthropogenic	Anthropogenic	1
Chloroethane	C <sub>2</sub> H <sub>5</sub> Cl	12 ppt	Anthropogenic	Anthropogenic	17
Trichloroethylene	C <sub>2</sub> HCl <sub>3</sub>	7.5 ppt	Anthropogenic	Red seaweeds	15
Dimethylsulfide	C <sub>2</sub> H <sub>6</sub> S	5–60 ppt	Biology	Marine plankton	16
F115	CClF <sub>5</sub>	4 ppt	Entirely anthropogenic	Anthropogenic	1

(continued)

TABLE A1. (CONTINUED)

<i>Compound name</i>	<i>Formula</i>	<i>Typical atmospheric concentration</i>	<i>Primary terrestrial atmospheric source</i>	<i>Example of biological production</i>	<i>Reference for biological production</i>
F116	C <sub>2</sub> F <sub>6</sub>	4 ppt	Entirely anthropogenic	Anthropogenic	1
F13	CClF <sub>3</sub>	3.3 ppt	Entirely anthropogenic	Anthropogenic	1
Methyl iodide	CH <sub>3</sub> I	2 ppt	Biology	Marine bacteria	12
F21	CHClF <sub>2</sub>	1.6 ppt	Entirely anthropogenic	Anthropogenic	1
Bromochlorodifluoromethane	CBrClF <sub>2</sub>	1.2 ppt	Entirely anthropogenic	Anthropogenic	1

Earth's atmospheric gases to the ppt level. All gases to the ppt level—regardless of the predominant production source and with the exception of the noble gases and fluorocarbons (anthropogenic)—are produced by life. Example references are provided for less common gases produced by life, mostly to a review rather than a complete reference list. Atmospheric list of gases and concentrations taken from Lodders and Fegley (1998), Seinfeld and Pandis (2006), and references therein. Note that the hydrocarbon “C<sub>2</sub>H<sub>6</sub> etc.” and “C<sub>6</sub>H<sub>6</sub> etc.” categories are rather broad, often locally measured rather than globally, and we address how many of them are actually made by life in our own analysis. The gases N<sub>2</sub>, O<sub>2</sub>, H<sub>2</sub>O, and CH<sub>4</sub> are produced by many organisms and can be found in many textbooks; they are therefore noted “1” (*e.g.*, Seinfeld and Pandis, 2006). The anthropogenically produced gas concentrations are noted “1” for similar reasons. Other references are as follows: (2) Hallenbeck and Benemann (2002); (3) Anderson and Levine (1986); (4) Ryter and Otterbein (2004); (5) Wentworth *et al.* (2003); (6) Paul and Pohnert (2011); (7) Monson *et al.* (2012); (8) Belay and Daniels (1987); (9) Pyysalo (1976); (10) Vitousek *et al.* (2013); (11) Belser (1979); (12) Gribble (2003); (13) Kuhn *et al.* (2000); (14) Nightingale *et al.* (1995); (15) Kladi *et al.* (2004); (16) Watts (2000); (17) Ballschmiter (2003); (18) Finster (2008); (19) Geng and Mu (2006).

TABLE A2. DATABASES FOR CHEMICALS

<i>Database</i>	<i>Literature reference (if any)</i>	<i>URL</i>	<i>Comment</i>
Dictionary of Natural Products	Dictionary of Natural Products, CRC Press, CD version	<a href="http://dnp.chemnetbase.com">http://dnp.chemnetbase.com</a>	
ZINC	John J. Irwin and Brian K. Shoichet. (2005) ZINC—A free database of commercially available compounds for virtual screening. <i>J Chem Inf Model</i> 45:177–182	<a href="http://zinc.docking.org/browse/catalogs/natural-products">http://zinc.docking.org/browse/catalogs/natural-products</a>	Natural products subset of the database. Only available as Web page
Sigma catalogue		<a href="http://www.sigmaaldrich.com">http://www.sigmaaldrich.com</a>	Commercial catalog, structure of all molecules ≤150 Da molecular weight (Bret Daniel, private communication, 2015) (Sigma-Aldrich)
Product catalogues from industrial gas suppliers		Compiled from Web catalogs of compounds provided by Air Liquide ( <a href="http://www.uk.airliquide.com">http://www.uk.airliquide.com</a> ) Linde ( <a href="http://www.linde.com">http://www.linde.com</a> ) British Oxygen Corp. ( <a href="http://www.boconline.co.uk">http://www.boconline.co.uk</a> ) Matheson Trigas ( <a href="http://www.mathesongas.com">http://www.mathesongas.com</a> )	
Reference books	Richard E. Lewis, Sr., editor. (2007) <i>Hawley's Condensed Chemical Dictionary</i> , 15 <sup>th</sup> ed., Wiley, Hoboken, NJ. Knovel Solvents—A Properties Database ( <a href="http://app.knovel.com">http://app.knovel.com</a> ) Chemical Properties Handbook ( <a href="http://app.knovel.com">http://app.knovel.com</a> ) Dictionary of Inorganic and Organometallic Compounds ( <a href="http://dioc.chemnetbase.com">http://dioc.chemnetbase.com</a> ) James Speight, editor. (2005)		From several chemical properties databases hosted by Knovel—apps. Knovel.com

(continued)

TABLE A2. (CONTINUED)

Database	Literature reference (if any)	URL	Comment
	<i>Lange's Handbook of Chemistry</i> , 16 <sup>th</sup> ed., McGraw-Hill, New York.		
	<i>Yaws' Handbook of Antoine Coefficients for Vapor Pressure</i> , 2 <sup>nd</sup> electronic edition ( <a href="http://app.knovel.com">http://app.knovel.com</a> )		
	Ernest Flick. (1999) <i>Industrial Solvents Handbook</i> , 5 <sup>th</sup> ed., Noyes Data Corporation, Westwood, NJ.		
	Merck Index ( <a href="https://www.rsc.org/merck-index">https://www.rsc.org/merck-index</a> )		
Pherobase		<a href="http://www.pherobase.com">http://www.pherobase.com</a>	

Chemical databases and other sources used for populating our list of stable and volatile chemicals.

## Appendix B

TABLE B1. SAMPLE TABLE FOR THE CNOPSH LIST

IUPAC name	Mol. formula	SMILES	Mol. weight (Da)	Boiling point (°C)	Life	Ref
N,N-Dimethylmethanamine	C <sub>3</sub> H <sub>9</sub> N	CN(C)C	59.11	4(e)	Y	1
Buta-2,3-dien-1-amine	C <sub>4</sub> H <sub>7</sub> N	NCC=C=C	69.1	88(p)	N	—
3-Buten-2-amine	C <sub>4</sub> H <sub>9</sub> N	C=CC(C)N	71.12	73(p)	N	—
2-Oxiranecarbaldehyde	C <sub>3</sub> H <sub>4</sub> O <sub>2</sub>	O=CC1OC1	72.06	96(p)	N	—
N-Ethylethanamine	C <sub>4</sub> H <sub>11</sub> N	CCNCC	73.14	55(e)	Y	2
Aminourea	CH <sub>5</sub> N <sub>3</sub> O	NNC(N)=O	75.07	210(p)	Y	3
Amino(hydroxyamino)methanol	CH <sub>6</sub> N <sub>2</sub> O <sub>2</sub>	OC(N)NO	78.07	221(p)	N	—
Cyclopropylmethanethiol	C <sub>4</sub> H <sub>8</sub> S	SCC1CC1	88.17	111(p)	N	—
1,3-Dithietane 1-oxide	C <sub>2</sub> H <sub>4</sub> OS <sub>2</sub>	O=S1CSC1	108.18	198(p)	N	—
Methyl hydrogen carbonotrithioate	C <sub>2</sub> H <sub>4</sub> S <sub>3</sub>	S=C(S)SC	124.25	200(p)	N	—

Columns are IUPAC name, molecular formula, SMILES description, molecular weight, boiling point (e=experimentally measured, p=predicted), production by life (Y/N), and references for molecules produced by life. References: (1) Kite and Hettterscheid (1997); (2) Smith and Meeuse (1966); (3) Van Poucke *et al.* (2011).

TABLE B2. SAMPLE TABLE FOR THE EXTENDED CNOPSH LIST

IUPAC name	Mol. formula	SMILES	Mol. weight (Da)	Boiling point (°C)	Life	Ref
Stibine	H <sub>3</sub> Sb	[SbH3]	124.78	-18(e)	Y	1
3-(Methylselanyl)-1-propene	C <sub>4</sub> H <sub>8</sub> Se	C[Se]CC=C	135.07	113(p)	N	—
Tetrahydroselenophene	C <sub>4</sub> H <sub>8</sub> Se	[Se]1CCCC1	135.07	128(p)	N	—
2-(Methylselanyl)-1-propene	C <sub>4</sub> H <sub>8</sub> Se	[Se](C(=C)C)C	135.07	107(p)	N	—
N,N-Dimethylselenoformamide	C <sub>3</sub> H <sub>7</sub> NSe	[Se]=CN(C)C	136.05	120(p)	N	—
Se-Methyl ethaneselenoate	C <sub>3</sub> H <sub>6</sub> OSe	O=C([Se]C)C	137.04	137(p)	N	—
Methylstibine	CH <sub>5</sub> Sb	C[SbH2]	138.81	41(e)	Y	2
[(Methylselanyl)sulfanyl]methane	C <sub>2</sub> H <sub>6</sub> SSe	C[Se]SC	141.09	135(p)	N	—
3-(Methyltellanyl)-1-propene	C <sub>4</sub> H <sub>8</sub> Te	[Te](CC=C)C	183.71	111(p)	N	—
Monomethyl bismuth hydride	CH <sub>5</sub> Bi	C[BiH2]	226.03	80(p)	Y	3

Columns are IUPAC name, molecular formula, SMILES description, molecular weight, boiling point (e=experimentally measured, p=predicted), production by life (Y/N), and references for molecules produced by life. References: (1) Michalke *et al.* (2000); (2) Wehmeier and Feldmann (2005); (3) Meyer *et al.* (2008).



TABLE B3. SAMPLE TABLE FOR THE HALOGENATED COMPOUNDS LIST

IUPAC name	Mol. formula	SMILES	Mol. weight (Da)	Boiling point (°C)	Life	Ref
Bromotrichloromethane	CBrCl <sub>3</sub>	C(Cl)(Cl)(Cl)Br	198.27	105(e)	N	—
Bromodiodomethane	CHBrI <sub>2</sub>	IC(I)Br	346.73	221.5(p)	Y	1
Chlorotribromomethane	CBr <sub>3</sub> Cl	C(Cl)(Br)(Br)Br	287.18	156(p)	N	—
Trichloroacetonitrile	C <sub>2</sub> Cl <sub>3</sub> N	N#CC(Cl)(Cl)Cl	144.38	84(e)	Y	2
1-Fluoropropane	C <sub>3</sub> H <sub>7</sub> F	CCCCF	62.09	-3(e)	N	—
Tetrafluoromethane	CF <sub>4</sub>	C(F)(F)(F)F	88	-130(e)	N	—
Cyanic chloride	CNCl	C(#N)Cl	61.47	13(e)	N	—
2-Bromo-1,1-dichloroethene	C <sub>2</sub> HBrCl <sub>2</sub>	C(=C(Cl)Cl)Br	175.84	109(p)	Y	3
Sulfur hexafluoride	SF <sub>6</sub>	FS(F)(F)(F)(F)F	146.06	-64(e)	N	—
Trichloroamine	NCl <sub>3</sub>	N(Cl)(Cl)Cl	120.37	71(e)	N	—

Columns are IUPAC name, molecular formula, SMILES description, molecular weight, boiling point (e=experimentally measured, p=predicted), production by life (Y/N), and references for molecules produced by life. References: (1) Gribble (2003); (2) Ballschmitter (2003); (3) Nightingale *et al.* (1995).

## References

- Abe, Y., Abe-Ouchi, A., Sleep, N.H., and Zahnle, K.J. (2011) Habitable zone limits for dry planets. *Astrobiology* 11:443–460.
- Alcolombri, U., Ben-Dor, S., Feldmesser, E., Levin, Y., Tawfik, D.S., and Vardi, A. (2015) Identification of the algal dimethyl sulfide-releasing enzyme: a missing link in the marine sulfur cycle. *Science* 348:1466–1469.
- Amo, L., Jansen, J.J., van Dam, N.M., Dicke, M., and Visser, M.E. (2013) Birds exploit herbivore-induced plant volatiles to locate herbivorous prey. *Ecol Lett* 16:1348–1355.
- Anderson, I.C. and Levine, J.S. (1986) Relative rates of nitric oxide and nitrous oxide production by nitrifiers, denitrifiers, and nitrate respirers. *Appl Environ Microbiol* 51: 938–945.
- Arnold, L. (2008) Earthshine observation of vegetation and implication for life detection on other planets. *Space Sci Rev* 135:323–333.
- Bains, W. (2004) Many chemistries could be used to build living systems. *Astrobiology* 4:137–167.
- Bains, W. and Seager, S. (2012) A combinatorial approach to biochemical space: description and application to the redox distribution of metabolism. *Astrobiology* 12:271–281.
- Bains, W., Xiao, Y., and Yu, C. (2015) Prediction of the maximum temperature for life based on the stability of metabolites to decomposition in water. *Life* 5:1054.
- Ballschmitter, K. (2003) Pattern and sources of naturally produced organohalogenes in the marine environment: biogenic formation of organohalogenes. *Chemosphere* 52:313–324.
- Bates, T.S., Kelly, K.C., and Johnson, J.E. (1993) Concentrations and fluxes of dissolved biogenic gases (DMS, CH<sub>4</sub>, CO, CO<sub>2</sub>) in the equatorial Pacific during the SAGA 3 experiment. *J Geophys Res: Atmospheres* 98:16969–16977.
- Beichman, C., Benneke, B., Knutson, H., Smith, R., Lagage, P.-O., Dressing, C., Latham, D., Lunine, J., Birkmann, S., Ferruit, P., Giardino, G., Kempton, E., Carey, S., Krick, J., Deroo, P.D., Mandell, A., Ressler, M.E., Shporer, A., Swain, M., Vasisth, G., Ricker, G., Bouwman, J., Crossfield, I., Greene, T., Howell, S., Christiansen, J., Ciardi, D., Clampin, M., Greenhouse, M., Sozzetti, A., Goudfrooij, P., Hines, D., Keyes, T., Lee, J., McCullough, P., Robberto, M., Stansberry, J., Valenti, J., Rieke, M., Rieke, G., Fortney, J., Bean, J., Kreidberg, L., Ehrenreich, D., Deming, D., Albert, L., Doyon, R., and Sing, D. (2014) Observations of transiting exoplanets with the James Webb Space Telescope (JWST). *Publ Astron Soc Pac* 126:1134–1173.
- Belay, N. and Daniels, L. (1987) Production of ethane, ethylene, and acetylene from halogenated hydrocarbons by methanogenic bacteria. *Appl Environ Microbiol* 53:1604–1610.
- Belser, L.W. (1979) Population ecology of nitrifying bacteria. *Annu Rev Microbiol* 33:309–333.
- Berg, J.M., Tymoczko, J.L., and Stryer, L. (2002) *Biochemistry*, 5<sup>th</sup> ed., W.H. Freeman, New York.
- Boublík, T., Fried, V., and Hála, E. (1973) *The Vapour Pressures of Pure Substances: Selected Values of the Temperature Dependence of the Vapour Pressures of Some Pure Substances in the Normal and Low Pressure Region*, Elsevier, Amsterdam.
- Brimblecombe, P. (2003) *The Global Sulfur Cycle*, Pergamon Press, Oxford, UK.
- Buckingham, J. (1993) *Dictionary of Natural Products*, Taylor & Francis.
- Budisa, N. and Schulze-Makuch, D. (2014) Supercritical carbon dioxide and its potential as a life-sustaining solvent in a planetary environment. *Life* 4:331–340.
- Burke, R.F. (1983) Biological activity of selenium. *Annu Rev Nutr* 3:53–70.
- Cabrita, M.T., Vale, C., and Rauter, A.P. (2010) Halogenated compounds from marine algae. *Mar Drugs* 8:2301–2317.
- Chopra, A. and Lineweaver, C.H. (2010) The stoichiometry of the essential elements of life [abstract 5561]. In *Astrobiology Science Conference 2010: Evolution and Life: Surviving Catastrophes and Extremes on Earth and Beyond*, Lunar and Planetary Institute, Houston.
- Christensen, P.R. and Pearl, J.C. (1997) Initial data from the Mars Global Surveyor thermal emission spectrometer experiment: observations of the Earth. *J Geophys Res* 102: 10875–10880.
- Cowan, D.A. (2004) The upper temperature for life – where do we draw the line? *Trends Microbiol* 12:58–60.
- Deisig, N., Dupuy, F., Anton, S., and Renou, M. (2014) Responses to pheromones in a complex odor world: sensory processing and behavior. *Insects* 5:399–422.
- Deming, D., Seager, S., Winn, J., Miller-Ricci, E., Clampin, M., Lindler, D., Greene, T., Charbonneau, D., Laughlin, G.,

- Ricker, G., Latham, D., and Ennico, K. (2009) Discovery and characterization of transiting super Earths using an all-sky transit survey and follow-up by the James Webb Space Telescope. *Publ Astron Soc Pac* 121:952–967.
- Des Marais, D.J., Harwit, M.O., Jucks, K.W., Kasting, J.F., Lin, D.N.C., Lunine, J.I., Schneider, J., Seager, S., Traub, W.A., and Woolf, N.J. (2002) Remote sensing of planetary properties and biosignatures on extrasolar terrestrial planets. *Astrobiology* 2:153–181.
- Domagal-Goldman, S.D., Meadows, V.S., Claire, M.W., and Kasting, J.F. (2011) Using biogenic sulfur gases as remotely detectable biosignatures on anoxic planets. *Astrobiology* 11:419–441.
- Domagal-Goldman, S.D., Segura, A., Claire, M.W., Robinson, T.D., and Meadows, V.S. (2014) Abiotic ozone and oxygen in atmospheres similar to prebiotic Earth. *Astrophys J* 792, doi:10.1088/0004-637X/792/2/90.
- Elkins-Tanton, L.T. and Seager, S. (2008) Ranges of atmospheric mass and composition of super-Earth exoplanets. *Astrophys J* 685, doi:10.1086/591433.
- El-Sayed, A.M. (2014) The Pherobase: database of pheromones and semiochemicals. Available online at <http://www.pherobase.com>
- Finster, K. (2008) Microbiological disproportionation of inorganic sulfur compounds. *J Sulphur Chem* 29:281–292.
- Fortes, A.D. (2000) Exobiological implications of a possible ammonia–water ocean inside Titan. *Icarus* 146:444–452.
- Fusetani, N. (2012) Marine natural products. In *Natural Products in Chemical Biology*, edited by N. Civjans, Wiley, Hoboken, NJ, pp 31–64.
- Gardner, J.P., Mather, J.C., Clampin, M., Doyon, R., Greenhouse, M.A., Hammel, H.B., Hutchings, J.B., Jakobsen, P., Lilly, S.J., Long, K.S., Lunine, J.I., McCaughrean, M.J., Mountain, M., Nella, J., Rieke, G.H., Rieke, M.J., Rix, H.-W., Smith, E.P., Sonneborn, G., Stiavelli, M., Stockman, H.S., Windhorst, R.A., and Wright, G.S. (2006) The James Webb Space Telescope. *Space Sci Rev* 123:485–606.
- Geng, C. and Mu, Y. (2006) Carbonyl sulfide and dimethyl sulfide exchange between trees and the atmosphere. *Atmos Environ* 40:1373–1383.
- Grant, W.D. (2004) Life at low water activity. *Philos Trans R Soc Lond B Biol Sci* 359:1249–1267.
- Greenwood, N.N. and Earnshaw, A. (1997) *Chemistry of the Elements*, 2<sup>nd</sup> ed., Butterworth-Heinemann, Oxford, UK.
- Gribble, G. (2002) Naturally occurring organofluorines. In *Organofluorines*, edited by A.H. Neilsons, Springer, Berlin, pp 121–136.
- Gribble, G.W. (2003) The diversity of naturally produced organohalogens. *Chemosphere* 52:289–297.
- Gu, J., Gui, Y., Chen, L., Yuan, G., Lu, H.-Z., and Xu, X. (2013) Use of natural products as chemical library for drug discovery and network pharmacology. *PLoS One* 8:e62839.
- Guinan, E.F., Ribas, I., and Harper, G.M. (2003) Far-ultraviolet emissions of the Sun in time: probing solar magnetic activity and effects on evolution of paleoplanetary atmospheres. *Astrophys J* 594, doi:10.1086/376859.
- Gunatilaka, A.L. (2012) Plant natural products. In *Natural Products in Chemical Biology*, edited by N. Civjans, Wiley, Hoboken, NJ, pp 3–29.
- Haas, J.R. (2010) The potential feasibility of chlorinic photosynthesis on exoplanets. *Astrobiology* 10:953–963.
- Hallenbeck, P.C. and Benemann, J.R. (2002) Biological hydrogen production; fundamentals and limiting processes. *Int J Hydrogen Energy* 27:1185–1193.
- Harman, C.E., Schwieterman, E.W., Schottelkotte, J.C., and Kasting, J.F. (2015) Abiotic O<sub>2</sub> levels on planets around F, G, K, and M stars: possible false positives for life? *Astrophys J* 812, doi:10.1088/0004-637X/812/2/137.
- Hartmans, S., Schmuckle, A., Cook, A.M., and Leisinger, T. (1986) Methyl chloride: naturally occurring toxicant and C-1 growth substrate. *J Gen Microbiol* 132:1139–1142.
- Hirner, A.V., Feldmann, J., Krupp, E., Grümping, R., Goguel, R., and Cullen, W.R. (1998) Metal(loid)organic compounds in geothermal gases and waters. *Org Geochem* 29:1765–1778.
- Howard, A.W. (2013) Observed properties of extrasolar planets. *Science* 340:572–576.
- Hu, R., Seager, S., and Bains, W. (2012) Photochemistry in terrestrial exoplanet atmospheres. I. Photochemistry model and benchmark cases. *Astrophys J* 761:166–194.
- Hu, R., Seager, S., and Bains, W. (2013) Photochemistry in terrestrial exoplanet atmospheres. II. H<sub>2</sub>S and SO<sub>2</sub> photochemistry in anoxic atmospheres. *Astrophys J* 769:6–19.
- Irwin, J.J. and Shoichet, B.K. (2005) ZINC – a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45:177–182.
- Jeans, J. (1930) *The Universe Around Us*, Cambridge University Press, Cambridge, UK.
- Johnston, A.W.B., Todd, J.D., Sun, L., Nikolaidou-Katsaridou, M.N., Curson, A.R.J., and Rogers, R. (2008) Molecular diversity of bacterial production of the climate-changing gas, dimethyl sulphide, a molecule that impinges on local and global symbioses. *J Exp Bot* 59:1059–1067.
- Kaltenegger, L. and Sasselov, D. (2010) Detecting planetary geochemical cycles on exoplanets: atmospheric signatures and the case of SO<sub>2</sub>. *Astrophys J* 708, doi:10.1088/0004-637X/708/2/1162.
- Kaltenegger, L., Traub, W.A., and Jucks, K.W. (2007) Spectral evolution of an Earth-like planet. *Astrophys J* 658, doi:10.1086/510996.
- Kashefi, K. and Lovley, D.R. (2003) Extending the upper temperature limit for life. *Science* 30:934.
- Katritzky, A.R., Allin, S.M., and Siskin, M. (1996) Aquathermolysis: reactions of organic compounds with superheated water. *Acc Chem Res* 29:399–406.
- Khalil, M.A.K. and Rasmussen, R.A. (1999) Atmospheric methyl chloride. *Atmos Environ* 33:1305–1321.
- Kiene, R.P., Linn, L.J., and Bruton, J.A. (2000) New and important roles for DMSP in marine microbial communities. *Journal of Sea Research* 43:209–224.
- Kite, G.C. and Hettterschied, W.L.A. (1997) Inflorescence odours of *Amorphophallus* and *Pseudodracontium* (Araceae). *Phytochemistry* 46:71–75.
- Kladi, M., Vagias, C., and Roussis, V. (2004) Volatile halogenated metabolites from marine red algae. *Phytochem Rev* 3: 337–366.
- Kopparapu, R.K., Ramirez, R.M., Kasting, J.F., Eymet, V., Robinson, T.D., Mahadevan, S., Terrien, R.C., Domagal-Goldman, S., Meadows, V., and Deshpande, R. (2013) Habitable zones around main-sequence stars: new estimates. *Astrophys J* 765, doi:10.1088/0004-637X/765/2/131.
- Kuchner, M.J. (2003) Volatile-rich Earth-mass planets in the habitable zone. *Astrophys J* 596, doi:10.1086/378397.
- Kuhn, U., Wolf, A., Gries, C., Nash, T.H., III, and Kesselmeier J. (2000) Field measurements on the exchange of carbonyl sulfide between lichens and the atmosphere. *Atmos Environ* 34:4867–4878.
- Lederberg, J. (1965) Signs of life: criterion-system of exobiology. *Nature* 207:9–13.

- Léger, A., Mariotti, J.M., Mennesson, B., Ollivier, M., Puget, J.L., Rouan, D., and Schneider, J. (1996) Could we search for primitive life on extrasolar planets in the near future? *Icarus* 123:249–255.
- Léger, A., Selsis, F., Sotin, C., Guillot, T., Despois, D., Mawet, D., Ollivier, M., Labèque, A., Valette, C., Brachet, F., Chazelas, B., and Lammer, H. (2004) A new family of planets? “Ocean-planets.” *Icarus* 169:499–504.
- Léger, A., Fontcave, M., Labeyrie, A., Samuel, B., Demangeon, O., and Valencia, D. (2011) Is the presence of oxygen on an exoplanet a reliable biosignature? *Astrobiology* 11:335–341.
- Lim, H., Kubota, K., Kobayashi, A., and Sugawara, F. (1998) Sulfur-containing compounds from *Scorodocarpus borneensis* and their antimicrobial activity. *Phytochemistry* 48: 787–790.
- Lodders, K. and Fegley, B. (1998) *The Planetary Scientists’s Companion*, Oxford University Press, New York.
- Lovelock, J.E. (1965) A physical basis for life detection experiments. *Nature* 207:568–570.
- Luger, R. and Barnes, R. (2015) Extreme water loss and abiotic O<sub>2</sub> buildup on planets throughout the habitable zones of M dwarfs. *Astrobiology* 15:119–143.
- Manley, S.L. (2002) Phytogenesis of halomethanes: a product of selection or a metabolic accident? *Biogeochemistry* 60: 163–180.
- McKay, C.P. and Smith, H.D. (2005) Possibilities for methanogenic life in liquid methane on the surface of Titan. *Icarus* 178:274–276.
- Meyer, J., Michalke, K., Kouril, T., and Hensel, R. (2008) Volatilisation of metals and metalloids: an inherent feature of methanoarchaea? *Syst Appl Microbiol* 31:81–87.
- Michalke, K., Wickenheiser, E.B., Mehring, M., Hirner, A.V., and Hensel, R. (2000) Production of volatile derivatives of metal(loid)s by microflora involved in anaerobic digestion of sewage sludge. *Appl Environ Microbiol* 66:2791–2796.
- Monson, R.K., Grote, R., Niinemets, Ü., and Schnitzler, J.-P. (2012) Modeling the isoprene emission rate from leaves. *New Phytol* 195:541–559.
- Montañés-Rodríguez, P., Pallé, E., Goode, P.R., and Martín-Torres, F.J. (2006) Vegetation signature in the observed globally integrated spectrum of Earth considering simultaneous cloud data: applications for extrasolar planets. *Astrophys J* 651, doi:10.1086/507694.
- Moss, S.J., Murphy, C.D., Hamilton, J.T.G., McRoberts, W.C., O’Hagan, D., Schaffrath, C., and Harper, D.B. (2000) Fluoroacetaldehyde: a precursor of both fluoroacetate and 4-fluorothreonine in *Streptomyces cattleya*. *Chem Commun* 2000: 2281–2282.
- Murphy, C.D., Schaffrath, C., and O’Hagan, D. (2003) Fluorinated natural products: the biosynthesis of fluoroacetate and 4-fluorothreonine in *Streptomyces cattleya*. *Chemosphere* 52: 455–461.
- Nadalig, T., Farhan, U.L., Haque, M., Roselli, S., Schaller, H., Bringel, F., and Vuilleumier, S. (2011) Detection and isolation of chloromethane-degrading bacteria from the *Arabidopsis thaliana* phyllosphere, and characterization of chloromethane utilization genes. *FEMS Microbiol Ecol* 77:438–448.
- Ni, X. and Hager, L.P. (1998) cDNA cloning of *Batis maritima* methyl chloride transferase and purification of the enzyme. *Proc Natl Acad Sci USA* 95:12866–12871.
- Nightingale, P.D., Malin, G., and Liss, P.S. (1995) Production of chloroform and other low molecular-weight halocarbons by some species of macroalgae. *Limnol Oceanogr* 40:680–689.
- O’Hagan, D. and Harper, D.B. (1999) Fluorine-containing natural products. *J Fluor Chem* 100:127–133.
- Otte, M.L., Wilson, G., Morris, J.T., and Moran, B.M. (2004) Dimethylsulphoniopropionate (DMS) and related compounds in higher plants. *J Exp Bot* 55:1919–1925.
- Palle, E., Osorio, M.R.Z., Barrena, R., Montanes-Rodriguez, P., and Martin, E.L. (2009) Earth’s transmission spectrum from lunar eclipse observations. *Nature* 459:814–816.
- Paul, C. and Pohnert, G. (2011) Production and role of volatile halogenated compounds from marine algae. *Nat Prod Rep* 28:186–195.
- Peters, R.A. and Shorthouse, M. (1967) Observations on the metabolism of fluoride in *Acacia georginae* and some other plants. *Nature* 216:80–81.
- Pierrehumbert, R. and Gaidos, E. (2011) Hydrogen greenhouse planets beyond the habitable zone. *Astrophys J* 734, doi:10.1088/2041-8205/734/1/L13.
- Pilcher, C.B. (2003) Biosignatures of early Earths. *Astrobiology* 3:471–486.
- Pomim, V.H. (2011) *Seaweed: Ecology, Nutrient Composition, and Medicinal Uses*, Nova Science Publishers, New York.
- Pyysalo, H. (1976) Identification of volatile compounds in seven edible fresh mushrooms. *Acta Chemica Scandinavica* 30b:235–244.
- Reymond, J.-L., Ruddigkeit, L., Blum, L., and van Deursen, R. (2012) The enumeration of chemical space. *Wiley Interdiscip Rev Comput Mol Sci* 2:717–733.
- Rhew, R.C., Østergaard, L., Saltzman, E.S., and Yanofsky, M.F. (2003) Genetic control of methyl halide production in *Arabidopsis*. *Curr Biol* 13:1809–1813.
- Rhew, R.C., Whelan, M.E., and Min, D.H. (2014) Large methyl halide emissions from south Texas salt marshes. *Biogeosciences* 11:6427–6434.
- Ricker, G.R., Winn, J.N., Vanderspek, R., Latham, D.W., Bakos, G.A., Bean, J.L., Berta-Thompson, Z.K., Brown, T.M., Buchhave, L., Butler, N.R., Butler, R.P., Chaplin, W.J., Charbonneau, D., Christensen-Dalsgaard, J., Clampin, M., Deming, D., Doty, J., De Lee, N., Dressing, C., Dunham, E.W., Endl, M., Fressin, F., Ge, J., Henning, T., Holman, M.J., Howard, A.W., Ida, S., Jenkins, J., Jernigan, G., Johnson, J.A., Kaltenegger, L., Kawai, N., Kjeldsen, H., Laughlin, G., Levine, A.M., Lin, D., Lissauer, J.J., MacQueen, P., Marcy, G., McCullough, P.R., Morton, T.D., Narita, N., Paegert, M., Palle, E., Pepe, F., Pepper, J., Quirrenbach, A., Rinehart, S.A., Sasselov, D., Sato, B., Seager, S., Sozzetti, A., Stassun, K.G., Sullivan, P., Szeggyorgyi, A., Torres, G., Udry, S., and Villaseñor, J. (2014) Transiting Exoplanet Survey Satellite (TESS). *Proc SPIE* 9143, doi:10.1117/12.2063489.
- Robinson, T.D., Meadows, V.S., Crisp, D., Deming, D., A’Hearn, M.F., Charbonneau, D., Livengood, T.A., Seager, S., Barry, R.K., Hearty, T., Hewagama, T., Lisse, C.M., McFadden, L.A., and Wellnitz, D.D. (2011) Earth as an extrasolar planet: Earth model validation using EPOXI Earth observations. *Astrobiology* 11:393–408.
- Roje, S. (2006) S-Adenosyl-L-methionine: beyond the universal methyl group donor. *Phytochemistry* 67:1686–1698.
- Rosenberg, E. (2008) Germanium: environmental occurrence, importance and speciation. *Reviews in Environmental Science and Bio/Technology* 8:29–57.
- Rothman, L.S., Rinsland, C.P., Goldman, A., Massie, S.T., Edwards, D.P., Flaud, J.M., Perrin, A., Camy-Peyret, C., Dana, V., Mandin, J.Y., Schroeder, J., McCann, A., Gamache, R.R., Wattson, R.B., Yoshino, K., Chance, K.V., Jucks, K.W., Brown, L.R., Nemtchinov, V., and Varanasi, P. (1998)

- The HITRAN molecular spectroscopic database and HAWKS (HITRAN Atmospheric Workstation): 1996 edition. *J Quant Spectrosc Radiat Transf* 60:665–710.
- Rugheimer, S., Kaltenecker, L., Zsom, A., Segura, A., and Sasselov, D. (2013) Spectral fingerprints of Earth-like planets around FGK stars. *Astrobiology* 13:251–269.
- Rugheimer, S., Kaltenecker, L., Segura, A., Linsky, J., and Mohanty, S. (2015) Effect of UV radiation on the spectral fingerprints of Earth-like planets orbiting M stars. *Astrophys J* 809, doi:10.1088/0004-637X/809/1/57.
- Ryter, S.W. and Otterbein, L.E. (2004) Carbon monoxide in biology and medicine. *BioEssays* 26:270–280.
- Sanchez, S., Guzman-Trampe, S., Avalos, M., Ruiz, B., Rodriguez-Sanoja, R., and Jiménez-Estrada, M. (2012) Bacterial natural products. In *Natural Products in Chemical Biology*, edited by N. Civjans, Wiley, Hoboken, NJ, pp 65–108.
- Scalo, J., Kaltenecker, L., Segura, A., Fridlund, M., Ribas, I., Kulikov, Y.N., Grenfell, J.L., Rauer, H., Odert, P., Leitzinger, M., Selsis, F., Khodachenko, M.L., Eiroa, C., Kasting, J., and Lammer, H. (2007) M stars as targets for terrestrial exoplanet searches and biosignature detection. *Astrobiology* 7:85–166.
- Schaefer, L. and Fegley, B. (2011) Atmospheric chemistry of Venus-like exoplanets. *Astrophys J* 729, doi:10.1088/0004-637X/729/1/6.
- Schindler, T.L. and Kasting, J.F. (2000) Synthetic spectra of simulated terrestrial atmospheres containing possible biomarker gases. *Icarus* 145:262–271.
- Schulze-Makuch, D. and Irwin, L.N. (2008) *Life in the Universe: Expectations and Constraints*, Springer-Verlag, Berlin.
- Seager, S. (2013) Exoplanet habitability. *Science* 340:577–581.
- Seager, S. and Bains, W. (2015) The search for signs of life on exoplanets at the interface of chemistry and planetary science. *Science Adv* 1, doi:10.1126/sciadv.1500047.
- Seager, S., Bains, W., and Hu, R. (2013a) Biosignature gases in H<sub>2</sub>-dominated atmospheres on rocky exoplanets. *Astrophys J* 777, doi:10.1088/0004-637X/777/2/95.
- Seager, S., Bains, W., and Hu, R. (2013b) A biomass-based model to estimate the plausibility of exoplanet biosignature gases. *Astrophys J* 775, doi:10.1088/0004-637X/775/2/104.
- Seager, S., Thomson, M., Turnbull, M., Sparks, W., Shaklan, S., Roberge, A., Kuchner, M., Kasdin, N.J., Domagal-Goldman, S., Cash, W., and others. (2015) *Exo-S: Starshade Probe-Class Exoplanet Direct Imaging Mission Concept*, Jet Propulsion Laboratory, Pasadena, CA.
- Segura, A., Krelove, K., Kasting, J.F., Sommerlatt, D., Meadows, V., Crisp, D., Cohen, M., and Mlawer, E. (2003) Ozone concentrations and ultraviolet fluxes on Earth-like planets around other stars. *Astrobiology* 3:689–708.
- Segura, A., Kasting, J.F., Meadows, V., Cohen, M., Scalo, J., Crisp, D., Butler, R.A.H., and Tinetti, G. (2005) Biosignatures from Earth-like planets around M dwarfs. *Astrobiology* 5: 706–725.
- Seinfeld, J.H. and Pandis, S.N. (2006) *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, 2<sup>nd</sup> ed., John Wiley and Sons, Hoboken, NJ.
- Selsis, F., Despois, D., and Parisot, J.-P. (2002) Signature of life on exoplanets: can Darwin produce false positive detections? *Astron Astrophys* 388:985–1003.
- Sharpe, S.W., Johnson, T.J., Sams, R.L., Chu, P.M., Rhoderick, G.C., and Johnson, P.A. (2004) Gas-phase databases for quantitative infrared spectroscopy. *Appl Spectrosc* 58: 1452–1461.
- Shkolnik, E.L. and Barman, T.S. (2014) HAZMAT. I. The evolution of far-UV and near-UV emission from early M stars. *Astron J* 148, doi:10.1088/0004-6256/148/4/64.
- Smith, B.N. and Meeuse, B.J.D. (1966) Production of volatile amines and skatole at anthesis in some arum lily species. *Plant Physiol* 41:343–347.
- Stapelfeldt, K., Belikov, R., Bryden, G., Cahoy, K., Chakrabarti, S., Marley, M., McElwain, M., Meadows, V., Serabyn, E., Trauger, J., and others. (2015) *Exo-C Imaging Nearby Worlds: Exoplanet Direct Imaging Coronagraph Probe Mission Study*, Jet Propulsion Laboratory, Pasadena, CA.
- Stefels, J. (2000) Physiological aspects of the production and conversion of DMSP in marine algae and higher plants. *Journal of Sea Research* 43:183–197.
- Steiger, S. and Stöckl, J. (2014) The role of sexual selection in the evolution of chemical signals in insects. *Insects* 5: 423–438.
- Stein, S.E. and Brown, R.L. (1994) Estimation of normal boiling points from group contributions. *J Chem Inform Comput Sci* 34:581–587.
- Sunda, W., Kieber, D.J., Kiene, R.P., and Huntsman, S. (2002) An antioxidant function for DMSP and DMS in marine algae. *Nature* 418:317–320.
- Tacke, R. (1999) Milestones in the biochemistry of silicon: from basic research to biotechnological applications. *Angew Chem Int Ed Engl* 38:3015–3018.
- Tait, F.K. and Moore, R.M. (1995) Methyl chloride (CH<sub>3</sub>Cl) production in phytoplankton cultures. *Limnol Oceanogr* 40: 189–195.
- Tennyson, J. and Yurchenko, S.N. (2012) ExoMol: molecular line lists for exoplanet and other atmospheres. *Mon Not R Astron Soc* 425:21–33.
- Thayer, J.S. (2002) Review: biological methylation of less-studied elements. *Appl Organomet Chem* 16:677–691.
- Tian, F., France, K., Linsky, J.L., Mauas, P.J.D., and Vieytes, M.C. (2014) High stellar FUV/NUV ratio and oxygen contents in the atmospheres of potentially habitable planets. *Earth Planet Sci Lett* 385:22–27.
- Tirindelli, R., Dibattista, M., Pifferi, S., and Menini, A. (2009) From pheromones to behavior. *Physiol Rev* 89:921–956.
- Todd, J.D., Rogers, R., Li, Y.G., Wexler, M., Bond, P.L., Sun, L., Curson, A.R.J., Malin, G., Steinke, M., and Johnston, A.W.B. (2007) Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. *Science* 315:666–669.
- Turnbull, M.C., Traub, W.A., Jucks, K.W., Woolf, N.J., Meyer, M.R., Gorlova, N., Skrutskie, M.F., and Wilson, J.C. (2006) Spectrum of a habitable world: Earthshine in the near-infrared. *Astrophys J* 644, doi:10.1086/503322.
- Vaillancourt, F.H., Yeh, E., Vosburg, D.A., Garneau-Tsodikova, S., and Walsh, C.T. (2006) Nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chem Rev* 106: 3364–3378.
- Van Poucke, C., Detavernier, C.L., Wille, M., Kwakman, J., Sorgeloos, P., and Van Peteghem, C. (2011) Investigation into the possible natural occurrence of semicarbazide in *Macrobrachium rosenbergii* prawns. *J Agric Food Chem* 59: 2107–2112.
- Visscher, P.T., Baumgartner, L.K., Buckley, D.H., Rogers, D.R., Hogan, M.E., Raleigh, C.D., Turk, K.A., and Des Marais, D.J. (2003) Dimethyl sulphide and methanethiol formation in microbial mats: potential pathways for biogenic signatures. *Environ Microbiol* 5:296–308.
- Vitousek, P.M., Menge, D.N.L., Reed, S.C., and Cleveland, C.C. (2013) Biological nitrogen fixation: rates, patterns and

- ecological controls in terrestrial ecosystems. *Philos Trans R Soc Lond B Biol Sci* 368, doi:10.1098/rstb.2013.0119.
- Wackett, L.P., Dodge, A.G., and Ellis, L.B. (2004) Microbial genomics and the periodic table. *Appl Environ Microbiol* 70:647–655.
- Watts, S.F. (2000) The mass budgets of carbonyl sulfide, dimethyl sulfide, carbon disulfide and hydrogen sulfide. *Atmos Environ* 34:761–779.
- Wehmeier, S. and Feldmann, J. (2005) Investigation into antimony mobility in sewage sludge fermentation. *J Environ Monit* 7:1194–1199.
- Wentworth, P., Nieva, J., Takeuchi, C., Galve, R., Wentworth, A.D., Dilley, R.B., DeLaria, G.A., Saven, A., Babior, B.M., Janda, K.D., Eschenmoser, A., and Lerner, R.A. (2003) Evidence for ozone formation in human atherosclerotic arteries. *Science* 302:1053–1056.
- Winn, J.N. and Fabrycky, D.C. (2015) The occurrence and architecture of exoplanetary systems. *Annu Rev Astron Astrophys* 53:409–447.
- Wordsworth, R. and Pierrehumbert, R. (2014) Abiotic oxygen-dominated atmospheres on terrestrial habitable zone planets. *Astrophys J* 785, doi:10.1088/2041-8205/785/2/L20.
- Wuosmaa, A.M. and Hager, L.P. (1990) Methyl chloride transferase: a carbocation route for biosynthesis of halometabolites. *Science* 249:160–162.
- Yoshida, Y., Wang, Y., and Zeng, T. (2004) A three-dimensional global model study of atmospheric methyl chloride budget and distributions. *J Geophys Res* 109, doi:10.1029/2004JD00495.
- Zhan, Z., Seager, S., Bains, W., Eckman, L., Gandhi, A., Lin, Z., Pajusalu, M., Petkowski, J., and Sha, L. (2015) Spectra for small volatile molecules for potential exoplanet biosignature gases. In *Extreme Solar Systems III*, American Astronomical Society, Washington, DC.
- Zsom, A., Seager, S., de Wit, J., and Stamenković, V. (2013) Toward the minimum inner edge distance of the habitable zone. *Astrophys J* 778, doi:10.1088/0004-637X/778/2/109.

Address correspondence to:

S. Seager

Department of Earth, Atmospheric, and Planetary Sciences  
Massachusetts Institute of Technology  
54-1626 77 Massachusetts Ave.  
Cambridge, MA 02139

E-mail: seager@mit.edu

Submitted 7 September 2015

Accepted 24 February 2016

#### Abbreviations Used

DMS = dimethyl sulfide  
DMSP = dimethylsulfoniopropionate  
JWST = James Webb Space Telescope  
STP = standard temperature and pressure