A Combinatorial Approach to Biochemical Space: Description and Application to the Redox Distribution of Metabolism

Citation: Bains, William, and Sara Seager. "A Combinatorial Approach to Biochemical Space: Description and Application to the Redox Distribution of Metabolism." Astrobiology 12.3 (2012): 271–281. Web.

As Published: http://dx.doi.org/10.1089/ast.2011.0718

Publisher: Mary Ann Liebert, Inc.

Persistent URL: http://hdl.handle.net/1721.1/72070

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

ASTROBIOLOGY
Volume 12, Number 3, 2012

DOI: 10.1089/ast.2011.0718

Hypothesis Article

A Combinatorial Approach to Biochemical Space: Description and Application to the Redox Distribution of Metabolism

William Bains[1,2] and Sara Seager[1,3]

Abstract

Redox chemistry is central to life on Earth. It is well known that life uses redox chemistry to capture energy from environmental chemical energy gradients. Here, we propose that a second use of redox chemistry, related to building biomass from environmental carbon, is equally important to life. We apply a method based on chemical structure to evaluate the redox range of different groups of terrestrial biochemicals, and find that they are consistently of intermediate redox range. We hypothesize the common intermediate range is related to the chemical space required for the selection of a consistent set of metabolites. We apply a computational method to show that the redox range of the chemical space shows the same restricted redox range as the biochemicals that are selected from that space. By contrast, the carbon from which life is composed is available in the environment only as fully oxidized or reduced species. We therefore argue that redox chemistry is essential to life for assembling biochemicals for biomass building. This biomass-building reason for life to require redox chemistry is in addition (and in contrast) to life's use of redox chemistry to capture energy. Life's use of redox chemistry for biomass capture will generate chemical by-products—that is, biosignature gases—that are not in redox equilibrium with life's environment. These potential biosignature gases may differ from energy-capture redox biosignatures. Key Words: Metabolism—Modeling studies—Redox—Biosignatures. Astrobiology 12, 271–281.

1. Introduction

One of the objectives of astrobiology is to identify general principles that govern the chemistry of life so that we can understand how those principles apply to extraterrestrial environments (Des Marais *et al.*, 2008). In the analysis of the chemistry of life, two classes of rules or principles are those that govern the structure of chemicals (structural chemistry) and those that govern the energetics of reactions (thermodynamics).

Thermodynamic analysis of chemical species and their transformation is commonly applied in astrobiology and particularly in the discussion of "biosignature" gasses that indicate the possible presence of life on a planet (Hitchcock and Lovelock, 1967; Lovelock and Kaplan, 1975). Thermodynamic analysis explores the energetics of proposed metabolic transformations, especially those that could capture the energy needed for life (Seager *et al.*, 2012). In contrast, structural analysis explores the nature of metabolism and its biosignatures from the standpoint of the structures of the chemicals involved. The nature of the energy needed to build those structures is not central to structural analysis.

This paper describes an approach to the systematic exploration of chemical structure in metabolism and applies this logic to the redox chemistry of terrestrial metabolism. We begin by introducing the concept of chemical space (Section 2.1) and define a measure of redox that can be easily applied to analyze chemical space (Section 3.1). We describe a computational algorithm that generates all possible structures in a chemical space (Section 3.2) and apply the algorithm to define the chemical space from which biochemicals are selected (Section 4). In Section 5, we discuss the implications of our findings about the redox range of biochemicals and the chemical space from which biochemicals are selected, in light of the environmental sources of the elements from which biochemicals are constructed.

2. Background

2.1. Chemical space: the phase space of possible chemical structures

"Chemical space" is the phase space of possible chemical structures (Lipinski and Hopkins, 2004). The definition of

[1]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
[2]Rufus Scientific Ltd., Melbourn, Royston, UK.
[3]Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

"possible chemical structure" depends on the atoms from which the chemicals are constructed and the environment in which the chemical is stable. In principle, there are an infinite number of different molecules that can be described; in practice, a molecular size limit must be used for a practical discussion.

Chemical space is large. It is estimated that there are between 10^{60} and 10^{120} organic "drug-like" molecules[1] (Bohacek *et al.,* 1996; Warr, 1997; Lipinski and Hopkins, 2004; Polley *et al.,* 2006; Turnbull and Swarbrick, 2009). The "size" of chemical space depends not only on the number of atoms in the molecules considered but also on how many different atoms and types of bonds between atoms can be used to build those molecules; these factors lead to the differences in the number of drug-like molecule estimates above.

The components of biochemistry are a very small set of molecules compared to all molecules that could exist. The number of biochemicals (nonpolymeric molecules generated by life) is probably no more than 10^5 (Pichersky and Gang, 2000). The selection of these 10^5 molecules from 10^{120} might be a result of historical accident or the result of functional constraint (Benner *et al.,* 2004). The accidental origin hypothesis suggests that specific events during the origin of life or its subsequent evolution may have yielded terrestrial biochemistry, and if life began anew different chance events would result in a different terrestrial biochemistry. In contrast, the functional constraint hypothesis suggests that there are functional reasons why any life emerging on an Earth-like planet will end up with a biochemistry similar to terrestrial biochemistry (Pace, 2001).

Ultimately, one would like to provide an explanation for collective properties of the chemistry of life, such as elemental composition, molecular size, bonding patterns, and so on. Our goal in this paper is to explain one feature of the chemical space of biochemistry—redox chemistry (see Section 2.2 for redox background). We are motivated to study redox chemistry both because it is widespread in metabolism and also because redox chemistry by-products are potential biosignature gases on exoplanets. Additionally, a computational study of the redox properties (relating redox to molecular structure as opposed to reaction energetics) of terrestrial biochemical molecules compared to all molecules is a tractable study for small molecules.

2.2. Redox chemistry as a parameter in chemical space

Redox (short for "reduction and oxidation") chemistry adds or removes electrons from an atom or molecule [see Metzler (1977), especially pp 172–181]. The removal of electrons from an atom or ion to form a more positive atom or ion is oxidation; addition of electrons to form a more negative atom or ion is reduction. For covalently bonded atoms, the oxidation state of each atom is calculated by adding up the number of bonds to more electronegative elements and subtracting bonds to more electropositive elements [for a more detailed introduction to electronegativity, electrode potential, and related topics, see Metzler (1977)]. For organic molecules, therefore, loss of a bond to an electronegative element, or gain of a bond to an electropositive element, is reduction. The reverse is oxidation. Thus, a molecule can be described as more or less reduced depending on the sum of the oxidation states of its nonhydrogen atoms, so reduction state can be a parameter in structurally defined chemical space.

Life uses redox for two principle functions:

- To execute metabolism—to convert environmental chemicals into biochemicals and to interconvert those biochemicals to build the complex molecules necessary for life.
- To capture energy—exploiting chemical energy gradients in the environment to generate and capture chemical energy.

Redox chemistry is universal in terrestrial life (*e.g.,* Nelson and Cox, 2009). Redox reactions are a major source of metabolic energy (*e.g.,* Lehninger, 1972). Because all life uses redox chemistry, the products of redox chemistry are often considered as strong candidates for "biosignatures" that can be used to remotely detect the presence of living organisms (Committee on the Origins and Evolution of Life, National Research Council, 2002). Hence, a strong motivation for considering whether redox chemistry should be present on planets beyond Earth is to direct the search for biosignatures (Seager *et al.,* 2012).

Redox chemistry is very widespread in the hundreds of biochemical reactions in primary metabolism [291 of 787 reactions in the core metabolism common to all terrestrial life are redox reactions (Fig. 1)]. But nearly all these reactions are related to the construction of the chemical components of life; only between 10 and 20 of the reactions in core metabolism directly capture chemical energy as ATP (the "energy currency" of the cell)[2]. This leads us to explore the extent to which redox chemistry is essential for synthesis of biomolecules, as distinct from generating energy for life. This paper addresses the importance of redox for the construction of the chemicals of biochemistry, as distinct from its role in energy generation, by systematically exploring the chemical space in which biochemistry resides and analyzing its redox properties.

3. Model

3.1. Rr: *a structural description of the redox state of a molecule*

Our aim is to explore the structural implications of redox state in the chemical space of chemicals. Exploring redox from a structural viewpoint requires a different approach from the usual, thermodynamic analysis of redox reactions. Redox chemistry is usually discussed in the context of oxidizing or reducing power, which is the ability of one

[1]A "drug-like" molecule is one which has a structure that an expert chemist would consider has potential to be absorbed through the gut, unlikely to be rapidly broken down by the body, unlikely to be toxic, etc., and with a molecular weight below 500 Da (Oprea *et al.,* 2001; Proudfoot, 2002; Hann and Oprea, 2004; Wunberg *et al.,* 2006).

[2]Of the redox reactions of primary metabolism, only those involved in oxidative phosphorylation (the electron transport chain reactions—between 10 and 20 reactions depending on how reducing equivalents enter the chain) and one substrate-level phosphorylation reaction in glycolysis generate ATP directly. The oxidation of glyceraldehyde-3-phosphate to generate 1,3-diphosphoglycerate both generates a high-energy phosphate and reduces NAD+ to NADH. Arguably, the ATP generation by pyruvate kinase and succinyl-Co-synthetase also represent coupling of oxidation in the previous metabolic step to energy capture.

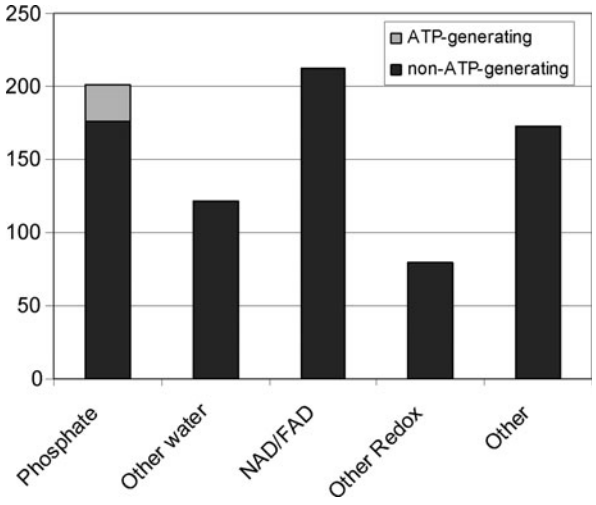


FIG. 1. Number of redox reactions in core metabolism. Number of reactions in "core" metabolism (intermediary metabolism) of heterotrophic, aerobic organisms that fall into different reaction categories. *y* axis: number of reactions. *x* axis: categories. "Phosphate" = involving hydrolysis of phosphate ester bonds. "Other water" = other reactions involving gain or loss of water from the molecule, without redox. "NAD/FAD" = redox reactions involving the carriers NAD, NADP, FAD. "Other redox" = redox reactions involving other donors or acceptors (including molecular oxygen). "Other" = reactions not falling into any of the previous four categories. Reactions are broken down into those that are directly involved in generating ATP (the "energy currency" of the cell) and those that are not involved in ATP generation. Data from the Roche metabolic map (Bairoch, 2000; ExPasy, 2010), collated by W.B.

molecule to oxidize or reduce another molecule, respectively. A common scale for comparing reducing power is the redox potential or electrode potential (see, *e.g.,* Nealson, 1997; Shock, 2009), a measure of the amount of energy that can be released by a redox reaction between two molecules. However, redox potential does not address the structure of the reacting molecules.

To compare the oxidation state of molecules of different size in relationship to the bonds between the atoms of those molecules, we define a "redox ratio" (*Rr*), which compares the bonding patterns of molecules by comparing the extent to which hydrogen atoms can be added to or removed from the nonhydrogen atoms of a molecule. The oxidation state of a molecule is closely tied to the amount of hydrogen in it. For nonmetals, and especially for carbon, reduction is equivalent to gaining hydrogen atoms. When no more hydrogen can be added to the atoms that make up the molecule, those atoms are fully reduced. The more hydrogen that can be added to the atoms in a molecule, the more oxidized a molecule is. This is a pragmatic, structure-based description of redox. We describe the *Redox Ratio (Rr)* of a molecule thus:

$$Rr = 1 - \frac{\sum S_A}{\sum S_H}$$

where S_A is the number of hydrogens bonded to each atom in the molecule and S_H is the maximum number of hydrogens that can be bonded to that atom. *Rr* is independent of the size and composition of the molecule and is independent of whether the atoms can have charges or not. Thus *Rr* for methane is 0, for formaldehyde 2/3, and for carbon dioxide 1. In the case of formaldehyde, *Rr* is calculated as follows. The number of hydrogens in formaldehyde is 2. The molecule contains 1 carbon atom, which can be joined to a maximum of 4 hydrogens, and 1 oxygen, which can be joined to a maximum of 2 hydrogen atoms. Thus $\sum S_H = 6$. Thus $Rr = 1 - \frac{2}{6} = \frac{4}{6} = \frac{2}{3}$. Joining any two molecules of $Rr = 0.67$ together, without losing or gaining any hydrogens in the process, will result in a molecule of $Rr = 0.67$.

We have introduced *Rr* as a simple, pragmatic description of the oxidation state of a molecule based on its bonding, not on its energetics. The redox ratio *Rr* is not meant to be profound, but convenient for such structural comparison.

This structural definition of redox in terms of possible bonds to hydrogen only makes sense for molecules (or their salts) built of elements that form bonds to hydrogen and for which forming a bond to hydrogen is reduction. This is roughly equivalent to elements with an Allred and Rochow electronegativity (Allred and Rochow, 1958) $> \sim 1.8$. These are B, Ga, C, Si, Ge, N, P, As, Sb, O, S, Se, Te, F, Cl, Br, and I. We emphasize again that we introduce *Rr* to describe the *structure* of molecules, *not* the thermodynamics of chemical reactions. Elements such as Fe, Mn, Ni, and so on are essential for life's use of redox as an energy source and for the enzymatic catalysis of some reactions, but few metabolites have metals as part of their covalent structure[3]. In this paper, we are concerned with the structure of the majority of chemicals in terrestrial biochemistry, so we limit our discussion to elements that are the most common *structural* components for the core set of biochemical molecules (C, N, O, H, S, and P) and describe them in structural terms, not in terms of thermodynamics. While the *Rr* of a series of similar compounds correlates roughly with their standard electrode potential (Fig. 2), the comparison is not really valid, as standard electrode potential relates to how compounds *react,* whereas *Rr* relates to their *structure*. Thus, a compound can have an *Rr* on its own, but its electrode potential relates to its *reduction*.

3.2. Computational combinatorics model

To explore the redox properties of chemical space, a way in which to describe all the molecules that can be made from a set of atoms is needed. The different chemistry of atoms means that not all combinations are possible. Simple combinatorics might say that there are 41 possible four-atom molecules to be made from C, N, and O (CCCC, CCCN, CCCO, etc.). But this does not count other topological possibilities such as branched molecules (C and N can link to three atoms but O cannot), rings (three- and four-member rings are possible), doubly bonded structures (all three atoms can form double bonds to another atom; only C and N can form triple bonds), or specifics of the chemistry such as the "rule-breaking" possibility of an amine oxide or the

[3]Many have cations as counter-ions to negatively charged groups such as carboxylate or phosphate, but cations *in this role* do not influence the chemistry of the metabolites significantly, with the possible exception of the complexation of polyphosphates with magnesium.

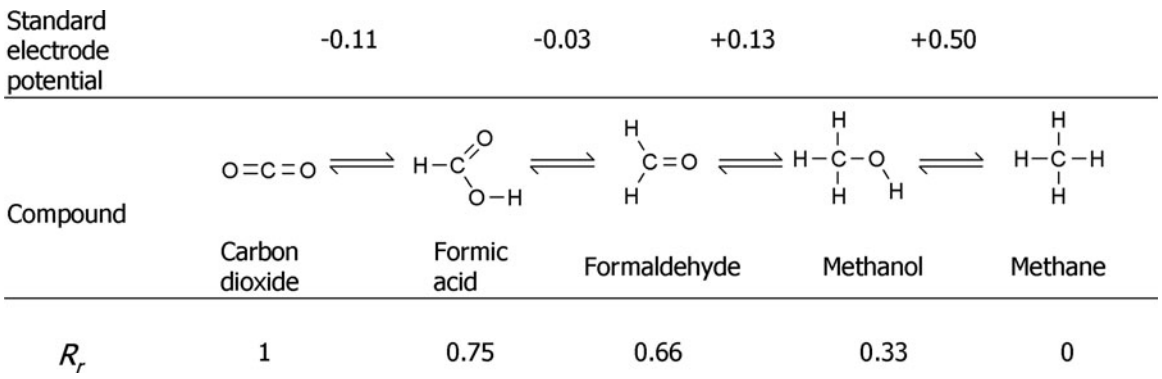


FIG. 2. Illustration of the relationship between standard electrode potential and *Rr*. Central panel: a series of one-carbon compounds. Top panel: standard electrode potential of half-cells generated by reduction of those compounds in pairs. Data from Bard *et al.* (1985). Lower panel: *Rr* for the individual compounds. Note that the standard electrode potential is a function of the *reaction* converting one compound to another, whereas *Rr* is a function of the structure of the compound.

impossibility of three oxygens in a chain. Counting molecules is harder than it appears.

One of us has developed a systematic algorithm for describing all the discrete, covalently bonded molecules that can be assembled from a set of atoms. The algorithm is implemented in the program COMBIMOL (Fortran, compiled for Windows). COMBIMOL can build molecules from C, H, and one or more of N, O, S, P, and Si. The program also offers the option to include or suppress compounds that contain groups likely to be reactive under physiological conditions, such as hydrazines, compounds likely to be reactive toward water, such as silanes, or compounds likely to be highly reducing, such as aldehydes or phosphines. These additional rules can be used to select the chemical space explored. The algorithm generates descriptions of molecules in the SMILES language (Weininger, 1988) (a simple, text-based molecular description syntax) under specific constraints of the rules of chemical bonding but without consideration of the energy requirement for their synthesis.

COMBIMOL generates all possible linear molecules from the specified set of atoms up to a maximum size by generating chains of carbons up to that size and then exhaustively substituting in all other atoms. Nonpermissible combinations, such as three oxygens in a row, are eliminated according to preprogrammed rules. The program next generates all possible ring systems that can be made by ring closing those chains into one or two rings, limited by a user-specified minimum size and again limited by bonding rules. The chains and rings are together termed "cores." Cores are then substituted into each other systematically by joining each pair of molecules using the SMILES syntax where a bond between two nonadjacent atoms is indicated by a number after those atoms, as illustrated in Fig. 3. Each new

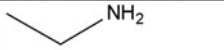
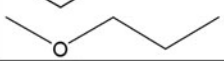
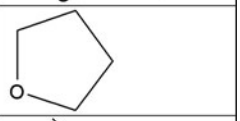
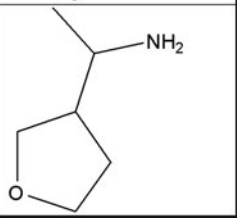
Molecule	SMILES	Structure diagram
ethylamine	NCC	
methyl-n-propyl ether	COCCC	
tetrahydrofuran	C1OCCC1	
2-(tetrahydrofuran-3-yl)ethanamine 3-(1-aminoethyl)tetrahydrofuran	C1OCC3C1.C3(C)N C1COCC13.C3(N)C C1OCC(C(N)C)C1	

FIG. 3. Illustration of SMILES-based molecular construction. First row: two simple linear molecules. COMBIMOL systematically generates all of these possible linear atom strings given a specified set of atoms and bonds. Second row: how one of the linear molecules is closed to a simple ring-structure molecule. (Note that the ring closure rules forbid three-member rings in this implementation, so the ethylamine is not converted to aziridine). Third row: one result of joining the two molecules in the first and second row together. First column: molecular name. Second column: SMILES string. Third column: structural diagram. Note that the compound described in the third row has at least two systematic names and many SMILES string descriptions that are compatible with the structure shown. The reader is referred to Weininger (1988) for details of the SMILES language.

molecule formed by joining two cores is then added to the list of cores for further substitution. This substitution cycle is continued until no new molecules within the target size range can be found.

Duplicates are avoided by (i) eliminating duplicates in the cores and (ii) sequentially substituting cores into each other. For example, if Core A is substituted into Core B, then Core B is not substituted into Core A. The core substitution test does not completely eliminate duplicates, so COMBIMOL will sometimes generate the same molecule twice. (Rigorously eliminating duplicates requires a structural comparison of every molecule generated with every other, requiring up to 10^{12} structural comparisons, which is impractical). A manual search of a sample of the data output suggests that $\sim$5% of the molecules generated are duplicates. We also note that the software cannot be formally proven to be completely exhaustive. Manual search, however, suggests that only a few percent of possible structures are not discovered.

The two biases of duplicates and missing molecules are opposite in direction. As our objective is to estimate the number of molecules in different classes, and not to exhaustively discover all structures, the likely error introduced is considered acceptable for this study.

The number of molecules generated by COMBIMOL is found to be approximated well by

$$\ln(M) = b \cdot N + c \cdot N^2$$

where M is the number of molecules, N is the number of atoms, and b and c depend on the types of atoms and the chemistry used to link them. (This is a purely empirical observation from generating 13.6 million structures by using 14 different combinations of atoms and bonds.) Values for b and c are given in Table 1, together with fractional errors in using the equation to predict the number of molecules found. Running time is roughly proportional to the number of molecules, generating $\sim 10^6$ structures/day on an IBM PC. Running speed therefore also limits the size of molecules that can be considered in practice.

4. Results

Our main finding is that both biochemicals and the chemical space from which biochemicals are selected are overwhelmingly concentrated in a narrow redox ratio (*Rr*) range. We find that the *Rr* of the chemicals of terrestrial life are generally concentrated in a relatively narrow range around 0.3; 74% have an *Rr* between 0.15 and 0.35, and almost no metabolites have an *Rr* of <0.1 or >0.5 (see Fig. 4). The narrow redox range distribution is almost identical for the four different groups of biochemicals we considered, even though these groups of biochemicals come from different areas of metabolism. Similarly, the *Rr* of the available molecules ("chemical space") from which biochemicals can be selected for life is also concentrated in a relatively narrow range around *Rr* of 0.25–0.4. (See Figs. 5 and 6.) The common narrow redox range implies causality, not a random historical chance. We speculate (in Section 5) the reason for the narrow redox range is that biochemistry requires a large chemical space, and the largest chemical space available is of intermediate redox range.

4.1. Redox state of biochemicals

We now describe the finding that biochemicals are concentrated in a narrow redox ratio (*Rr*) range. Seventy-four percent have an *Rr* between 0.15 and 0.35. We have compiled four sets of metabolites to represent the diversity of terrestrial metabolism (see Table 2). One set of metabolites represents the core of common metabolism on Earth, and the three other sets of metabolites represent different specializations of metabolism in major groups of organisms.

The "core" metabolism is the network of chemical reactions that are central to capturing energy and making the precursors of proteins, fats, sugars, and nucleic acids in all cells. We pragmatically define the "core" set as the molecules shown on the commonly used metabolic pathway charts (Bairoch, 2000; ExPasy, 2010); there are around 730 metabolites in this "core" set. The pathways have been derived primarily from heterotrophic organisms, but much applies to

Table 1. Formulae for Number of Molecules Generated by COMBIMOL, for Different Chemical Classes

Atom/bond combination	*Maximum* N	*log*(M_7)	b	c	*Error*
C (carbons only)	11	5.298	0.03240	0.102192	0.112
C O (includes carbonyls)	11	7.172	0.37933	0.089797	0.163
C N (does not include hydrazines)	10	8.072	0.53093	0.087250	0.118
C N (with hydrazines)	10	8.494	0.62828	0.082423	0.094
C N O (no hydrazines)	9	9.184	0.64412	0.093380	0.097
C N O (with hydrazines)	9	9.651	0.75297	0.087963	0.085
C N S (with hydrazines, S as thiol and thioether only)	9	9.515	0.84070	0.073433	0.065
C N S (as above, with sulfones, sulfoxides, sulfonamides, etc.)	9	9.168	0.76447	0.077342	0.078
C N S O P (as above, with phosphates and oxygen groups; broadly equivalent to terrestrial biochemistry)	8	10.432	0.89687	0.084038	0.071
C N S O Si (Si-H and Si-N bonds allowed, S-H bonds allowed in all contexts, S+ groups allowed)	9	12.210	1.13579	0.086238	0.067
C N O Si (Si only bonded to C, O)	8	8.026	0.39080	0.107438	0.051
C N O S P Si (all bonds and groups stable in inert solvents allowed)	8	12.899	1.27027	0.082331	0.047

Column 1: atoms and bonds used. Column 2: largest size of molecules generated by these COMBIMOL runs (limited by running time). Column 3: LN(number of molecules with 7 atoms in) for comparison. Columns 3 through 5: parameters in the equation $M = b \cdot N + c \cdot N^2$, where M is the number of molecules and N is the number of atoms used to generate those molecules. Column 6: RMS difference between predicted log(M) and log(M) found by counting molecules with COMBIMOL.

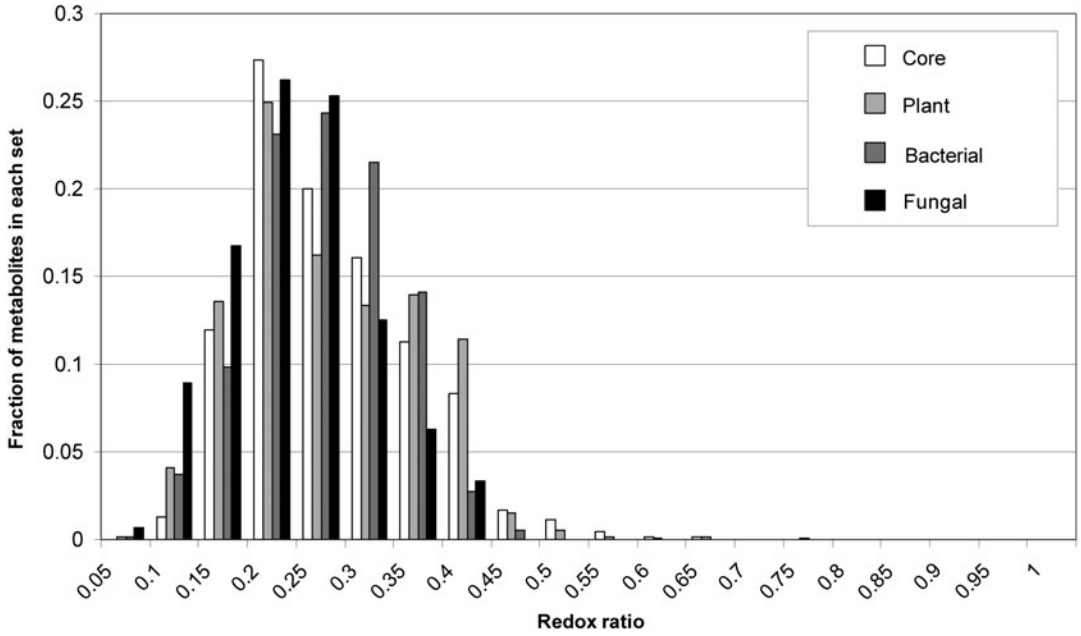


FIG. 4. Comparative oxidation state of metabolites. Shown is the Redox ratio (*Rr*) (*x* axis, binned in 0.05 increments) of four sets of metabolites: those from the core of metabolism common to all terrestrial life and three specific to plants, bacteria, and fungi. *y* axis: fraction of metabolites in each database falling into each *Rr* bin. The *Rr* is the extent to which the atoms in the molecules are fully bonded to hydrogen (*i.e.,* fully reduced), ranging from 0 (fully bonded to hydrogen, *e.g.,* CH_4 or NH_3) to 1 (not bonded to hydrogen at all, *e.g.,* CO_2) (Section 3.1). Seventy-four percent of metabolites have *Rr* between 0.15 and 0.35.

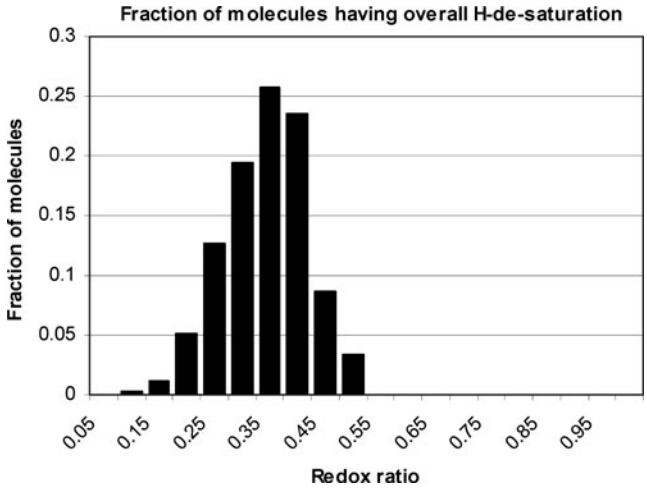


FIG. 5. Redox ratio of "biochemical space." The possible chemical diversity that can be generated from different combinations of atoms, as a function of the oxidation state of the molecules concerned. *y* axis: fraction of molecules of three through nine non-H atoms (total number of molecules 1,686,196). *x* axis: *Rr* values of the molecules. Molecules are built from C, N, O, S, and P with similar limits to bonding as seen in terrestrial biochemistry. The core of terrestrial biochemistry contains very few compounds containing O-O or N-N bonds, and S-S bonds are almost exclusively confined to disulfide bridges in proteins, so these were excluded. Phosphorus is included in molecules solely in its fully oxidized form (as phosphate groups).

chemotrophs and photosynthesizers as well. Many of the reactions in "core metabolism" are common to all terrestrial biochemistry and are called "intermediary metabolism" or "primary metabolism" (Alberts *et al.,* 1983). However, "core metabolism" also includes reactions important to chemotrophic and photosynthetic energy capture.

The more specialized metabolite sets are made by plants, fungi, and bacteria for specific functions such as defense against attackers or to act as hormones or signaling molecules. These more specialized sets of metabolites were included in our analysis to capture the diversity of chemistry that life develops. Such specialized metabolic pathways are called "secondary metabolism" and their products "secondary metabolites" because of their specialist nature. All duplicates within and between data sets were removed so that none of the molecules in the plant, fungal, or bacterial sets were core metabolites.

The distributions of *Rr* of the four sets of metabolites are shown in Fig. 4. Metabolites are almost always in a range of intermediate oxidation states (that is to say, the metabolites are mostly but not completely reduced).

We emphasize two findings about the *Rr* distribution of Earth-life metabolites. First, the majority of the molecules of metabolism fall in a relatively narrow range of *Rr*. Seventy-four percent fall between $Rr = 0.15$ and $Rr = 0.35$, and 95% between $Rr = 0.1$ and $Rr = 0.45$. Second, the relatively narrow range holds for all the four sets of biochemicals, even though these biochemicals come from areas of metabolism that have different functions (Fig. 4). While biochemists familiar with intermediary metabolism will not find the redox distribution of "core" metabolism very surprising, it is important that we establish the distribution of *Rr* shown in Fig. 4 before proceeding to discuss potential explanations for that distribution.

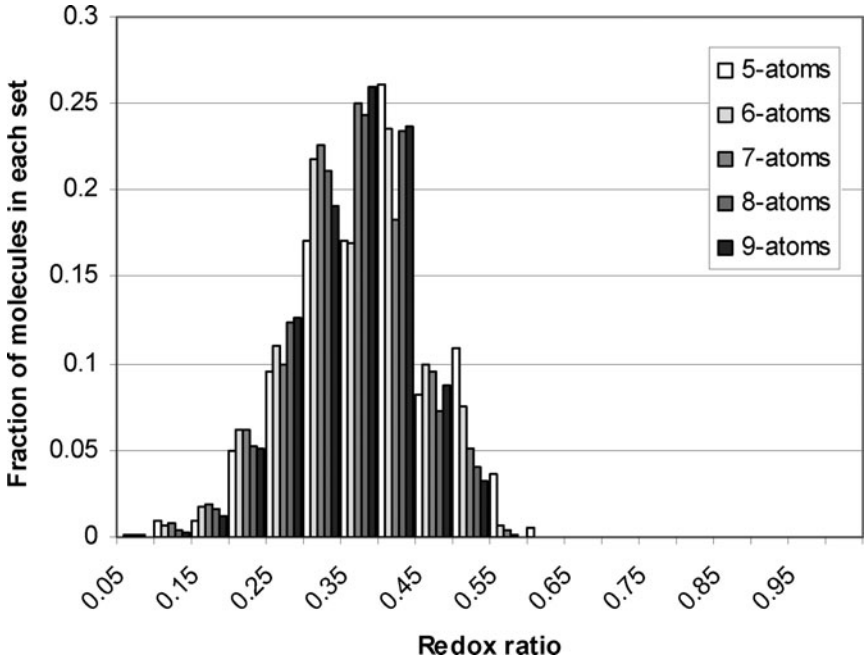


FIG. 6. Chemical diversity and redox by molecular size. The same analysis of the frequency distribution of *Rr* values for "biochemical space" as in Fig. 5 but plotted for molecules of different size. *x* axis: *Rr* value. *y* axis: fraction of molecules in "biochemical space" of each size class possessing a specific *Rr* value. Shown are the distributions for molecules containing five, six, seven, eight, and nine non-H atoms. The distribution is essentially the same for different molecule sizes. (There are 525 five-atom molecules, 3,412 six-atom molecules, 23,312 seven-atom molecules, 186,192 eight-atom molecules, and 1,472,755 nine-atom molecules.) This figure demonstrates that the *Rr* of chemical space is not dependent on molecular size.

We also note that it is rare for discussions of the properties of the chemicals of life to include any discussion of redox beyond its relevance to energy production.

4.2. Redox space from which biochemicals are selected

We now turn to the redox distribution of the chemical space of possible molecules from which biochemicals are selected ("biochemical space") in order to explain the very consistent redox range of terrestrial biochemicals (described in Section 4.1).

We want to calculate *Rr* for "biochemical space." We define biochemical space as all molecules containing C, N, O, S, P, H in bonding patterns found at least once in core metabolism. Applying COMBIMOL to generate all possible small molecules with these constraints, and evaluating *Rr* of those small molecules, we find an *Rr* distribution similar to that of terrestrial biochemicals (Fig. 5). It is clear that the large majority of chemicals have an intermediate oxidation state. Quantitatively, the *Rr* of 95% all possible small molecules lies between 0.15 and 0.45 (as compared to the *Rr* range of 95% of the pooled metabolites analyzed in this study lying between 0.1 and 0.4); one 68% lie between 0.25 and 0.4 (74% of metabolites lie between 0.15 and 0.35) (Figs. 5 and 6). Only 0.5% have an *Rr* of less than 0.1 or more than 0.5. As expected from the definition of *Rr*, the shape of the *Rr* range distribution is independent of size (Fig. 6), so, as *Rr* is conserved when molecules are added together, we can infer that this *Rr* distribution also holds for much larger molecules. We have not explored the *Rr* of larger molecules because of the computational limits implied by Table 1 (there are expected to be $\sim 10^{87}$ molecules of the same size as NADH, for example, which are impractical to enumerate).

Table 2. Sources of the Chemical Structures of Metabolites Used in This Study

Data set	*Source organisms*	*Source of chemical structures*	*URL*	*Reference*	*Number of compounds*
Core	All	Roche metabolic map (entered and curated by W.B.)	http://expasy.org/cgi-bin/show_thumbnails.pl	ExPasy, 2010	731
Plant	Plants (secondary metabolites)	Plant Metabolic Network "PlantCyc"	ftp://ftp.plantcyc.org/Pathways/plantcyc_compounds.20091014	Karp *et al.*, 2010	2072
Bacterial	*Streptomyces*, *Pseudomonas*, and *Bacillus* spp.	KNapSac	http://kanaya.naist.jp/KNApSAcK	Shinbo *et al.*, 2006	776
Fungal	Range of fungi	Nielsen and Smedsgaard, 2003 (entered and curated by W.B.)	n/a	Nielsen and Smedsgaard, 2003	781

The distribution of *Rr* in the metabolites of terrestrial biochemistry lies within the expected range of maximum possible diversity that can be created from the atoms and bonds that Earth life uses (compare Fig. 4 and Fig. 5). The concept of *Rr* does not apply easily to macromolecules. However, we do not consider this a significant limitation of our analysis. The major terrestrial biological macromolecules—proteins, carbohydrates, and nucleic acids—are made by dehydration-condensation of small metabolites. Dehydration of two molecules to form one molecule and a water molecule does not alter the combined *Rr* of the molecules, and metabolism must make the small metabolites before they can be polymerized. Recall that we defined *Rr* to be a ratio that is a measure of the bonding patterns between atoms in a molecule (Section 3.1).

We infer that terrestrial biochemicals have an intermediate redox state because (i) the many constraints on biochemistry mean that life must have many chemical options from which to select, that is, biochemical space must be large, and (ii) intermediate redox states, of *Rr* between 0.2 and 0.4, provide the majority of chemical space formed from the atoms and bonds available to life. Terrestrial biochemicals are slightly more reduced than the average of the chemical space from which they are selected. From this analysis, we cannot say whether this offset of the distribution has a deep meaning.

It is difficult to identify the constraints on biochemicals. Apart from general constraints on solubility and stability, a biochemical has to "fit" with the rest of metabolism, so its two- and three-dimensional structure is constrained to be compatible with a very large number of interactions, many of which we do not know about. However, we know that the constraints are very hard to satisfy, as illustrated by the high failure rate of the drug discovery process in the pharmaceutical industry. Most of the possible chemicals made up of C, N, O, S, or P are not compatible with biochemistry—they do not have a useful function, or they interfere with the cell's existing function. New drugs are discovered (in part) by screening large collections of chemicals for their ability to affect a particular protein that is believed to be important to a disease process. Chemicals that are successful in such a screen are termed "hits." Pharmaceutical chemists then make many analogues of the "hits," molecules that did not exist in the original collections, to find one that both has the pharmacological function required and that does not adversely affect other body processes. This is a notoriously failure-prone process. Only between 1 in 10,000 and 1 in 50,000 of these "hits" leads to analogues that make it to become candidates for sale to the public as drugs (Oprea and Marshall, 1998; Entzeroth, 2003; Keseru and Makara, 2006; Polley *et al.*, 2006). Many of the failed "hits" are in some way incompatible with the other functions of our cells, generating unwanted or toxic effects. While this is not a quantitative measure of constraint, the statistics of drug discovery illustrate the problem.

The narrow distribution of *Rr* in chemical space is not a unique feature of biochemical space. For any molecules made of atoms for which *Rr* can be defined (see Section 3.1), we expect there will be few with extreme *Rr* values and many with intermediate values. The range of *Rr* in chemical space where the maximum diversity lies may differ for different atom combinations. Initial exploration of variant chemistries with COMBIMOL (specifically, chemistry with C and different combinations of O, N, S, P, and Si) has confirmed that *Rr* restriction in chemical space is likely to apply to any covalent chemistry.

5. Discussion

We have evaluated the redox range of chemicals related to terrestrial biochemicals. We have defined a pragmatic measure of redox state—*Rr*—that is relevant to the structure of the molecule. We find that several sets of metabolites, derived from different organisms and with different functions, have a surprisingly similar, narrow distribution of *Rr*, with 74% of metabolites falling between $Rr = 0.15$ and $Rr = 0.35$, and 95% between $Rr = 0.1$ and $Rr = 0.45$. To explore one explanation for the similar distribution of *Rr* between groups of biochemicals that have different functional and evolutionary origins, we also explored the *Rr* of the chemical space from which biochemicals are selected. We found that the *Rr* of the chemical space of possible biochemicals is similar to that of actual biochemicals, with 68% lying between 0.25 and 0.4, and 95% lying between 0.15 and 0.45.

5.1. A restricted intermediate oxidation state for biochemical space

For life to be so consistent in the *Rr* states of its metabolites suggests that there is a reason for biochemicals to have a restricted distribution of intermediate oxidation states. This is especially true given the diverse gene sequences, shapes, sizes, life histories, and habitat chemistry among the animals, plants, fungi, and bacteria whose metabolites are represented in the sets that we used in our analysis in Fig. 4 (listed in Table 2). We hypothesize that the reason that different sets of actual biochemicals all have a restricted *Rr* between 0.1 and 0.45 is related to the function of those biochemicals within a living organism.

To function in a cell, a biochemical must carry out a number of specific tasks and not interfere with all the other machinery of a cell. We do not know what the function of each biochemical is or the exact nature of the functional constraints. However, we know from experience that the probability that a chemical chosen at random from "biochemical space" can fulfill those constraints is small. To maximize the chance of finding a compound that fulfills the necessary functional constraints, life needs to have many possible chemicals to chose from, that is, a large chemical space. The chemical space of possible biochemicals is overwhelmingly concentrated in the *Rr* range 0.1–0.45. Thus, it is from this *Rr* range that life selects its biochemicals.

An alternate hypothesis for the origin of the intermediate *Rr* state of the biochemicals of life is that this is a result of the chemistry that gave rise to life. Morowitz and Smith (Morowitz *et al.*, 2000; Smith and Morowitz, 2004), for example, argued that the chemistry of some of the most ancient, central reactions in primary metabolism derives from the redox chemistry of prebiotic chemistry and, hence, is likely to be universal. Martin and Russell (2007) argued similarly that the nature of core metabolism (in their case thioester chemistry) derives from the ancient, prebiotic chemistry from which life arose and, hence, is chemically inevitable.

We argue, however, that the constraints on redox chemistry at the origin of life are not the reason for the distribution

of oxidation states that is shown in Fig. 4. The ancestral, original biochemistry of life 3.8 billion years ago may explain the *Rr* of some aspects of core metabolism but cannot explain why secondary metabolites share the same *Rr* distribution with core metabolism. Secondary metabolites have specific functions that are related to their producer's current ecology and must have evolved relatively recently. An ancient origin, frozen in modern biochemistry, cannot explain the *Rr* of secondary metabolism.

5.2. *Environmental* Rr *versus biochemical* Rr

It is interesting to consider the *Rr* range of environmental chemicals compared to the *Rr* range of biochemicals. Carbon is the most common nonhydrogen atom in terrestrial biochemicals, and 90% of the bonds in the "core" metabolism are to carbon atoms. The carbon in biochemicals is itself of intermediate oxidation state (Fig. 7).

The environment of terrestrial planets (and large, atmosphere-holding moons such as Titan) is expected only to provide carbon in fully oxidized or fully reduced states. In a reduced environment carbon is most stable as methane, in an oxidizing environment as carbon dioxide. Life in such environments will therefore have to oxidize methane or reduce carbon dioxide in order to build the chemical diversity necessary for life, regardless of the specifics of biochemistry. In doing so, life must reduce or oxidize an environmental chemical and generate a by-product that does not contain carbon. We note that smaller bodies, such as meteorites or dust particles, are likely to harbor carbon in intermediate redox states, because loss of volatiles from those bodies will favor the spontaneous formation of compounds with little hydrogen or oxygen in them. However, meteorites or interplanetary dust is an unlikely environment for life to originate.

5.3. *Redox for biochemical structure as opposed to redox chemistry for energy generation*

We emphasize that our use of *Rr* as a description of redox state is *unrelated* to the analysis of redox gradients as an energy source. However, we cannot ignore the energetic implications of life's need to build molecules of intermediate *Rr*. Life's creation of compounds of an intermediate *Rr* by driving the oxidation or reduction of environmental compounds creates a redox couple. The creation of a reduced biomolecule and an oxidized environmental molecule is the production of a high-energy redox couple from a low-energy starting point. Creation of such a couple requires the input of energy.

Creation of a redox couple requires chemical energy. Obviously, life cannot use the construction of an intermediate redox compound as a *source* of energy—the creation of an intermediate redox compound synthesis is a *sink* or *use* of energy captured elsewhere. Life therefore must capture chemical energy to use in the reactions that capture carbon and turn it into molecules of intermediate *Rr*.

We also note that we are not implying that life reduces carbon from CO_2 in a single step. This is an analysis of *structure,* not *mechanism.* Life can perform reactions that

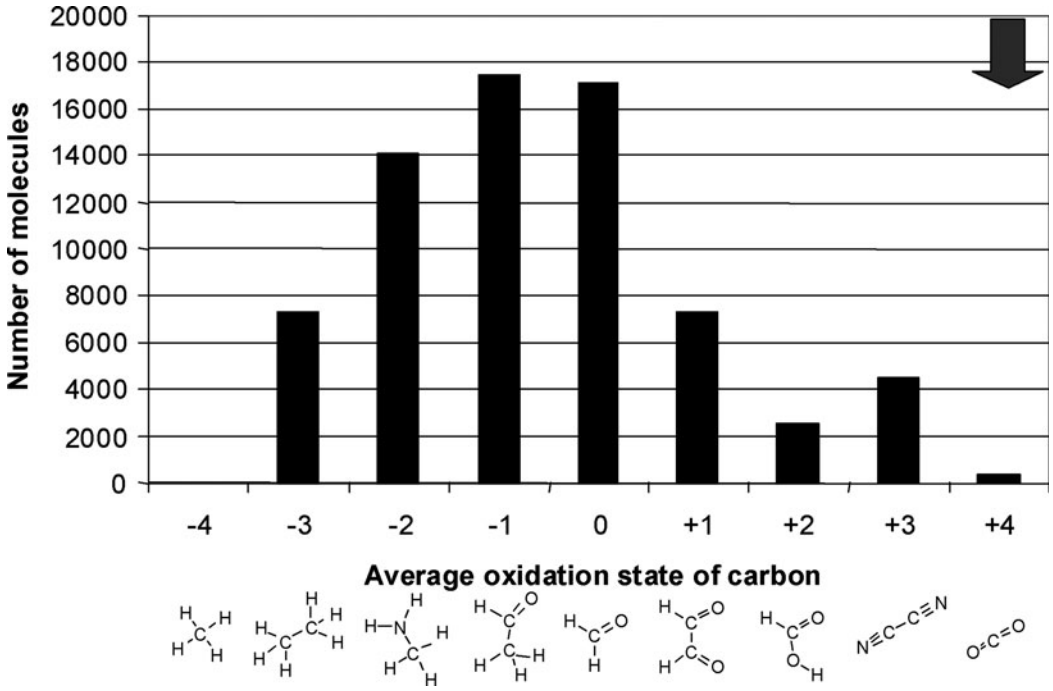


FIG. 7. Distribution of the oxidation state of the carbon atoms in terrestrial biochemicals and in the nonliving terrestrial environment. We calculate the average oxidation state (rounded to the nearest whole number) of the carbon atoms in each molecule in our collection of biochemicals. *x* axis: the average oxidation state of the carbon atoms. *y* axis: the number of molecules in the sets of metabolites analyzed in this paper containing carbon with that average oxidation state. To illustrate what the oxidation state means in chemical terms, a small organic molecule is drawn under each oxidation state in which the carbon has exactly that oxidation state. The gray arrow shows the oxidation state of carbon in the crust and the atmosphere of Earth, where carbon (other than that representing material generated by life such as coal and oil) is overwhelmingly present as carbonate or CO_2.

require substantial energy by breaking them down into several steps each of which requires less energy. Examples are the two-photon generation of NADPH and O_2 from H_2O and NADP in photosynthesis, the multiproton synthesis of ATP in the mitochondrial ATPase, even the low-error (and hence high-energy) synthesis of eukaryotic nuclear DNA by proof-reading DNA polymerases. Our argument is about the need for *structure* and not the energetics of how that structure is created.

Redox reactions can also be a source of energy for life to capture. Inevitably, some organisms may use redox both for energy generation and to build biochemicals. Some organisms may use the same redox reactions for these two, different purposes. Does it matter, then, why life uses redox chemistry? While not wishing to stray into teleology, we believe that it is important to dissect why metabolism is built as it is, so that we might seek to understand how it would operate in environments different from our own and, hence, inform our search for life on other worlds.

5.4. Biosignatures and other worlds

Our finding that an intermediate *Rr* is an apparently universal feature of terrestrial biochemicals has implications for the search for life. As noted in the introduction (Section 2.2), life uses redox biochemistry to capture energy and (using that energy) to capture and transform biomass. In this paper, we hypothesize that the biochemicals that are used to build biomass and, hence, the biomass itself must be of intermediate *Rr*. This conclusion is independent of the specifics of the metabolic pathways involved. The requirement that life have a substantial chemical space from which to select molecules that satisfy functional criteria means that those molecules will have to be of intermediate *Rr*. The requirement for a substantial chemical space does not, however, put other constraints on the nature of life.

The constraint that biochemicals should have intermediate *Rr* suggests that primary producers in an ecology will need to perform redox chemistry to build biomass, because the environmental sources of biomass are not of intermediate *Rr* (Section 5.3). This redox chemistry may have a byproduct gas, detectable by remote space telescopes or planetary orbiters, which could be a biosignature. It will be useful to consider both the energy-generating and the biomass-building role of redox when predicting the potential biosignature gases that could be generated by life in environments different from Earth's.

None of the analysis in this paper is specific to the geochemistry of Earth or even to terrestrial biochemistry. Initial explorations with COMBIMOL suggest that other spaces of carbon-based chemistry, such as the "ammonochemistry" suggested for Europa (Raulin *et al.*, 1995; Bains, 2004), also have a chemical-space *Rr* almost entirely between 0.1 and 0.5. We would expect carbon to be present as methane or as carbon dioxide in any terrestrial planetary environment; therefore, we would expect life to require energy-consuming redox chemistry to build biomass in those environments. This preliminary result requires more confirmation, and the result needs to be integrated with the geochemistry of the environment concerned and the thermodynamics of potential energy-yielding reactions in those environments. However, the analysis of the *Rr* of terrestrial metabolism described in this paper illustrates that exploring chemical space in a quantitative model-based way has interest beyond this paper's application of the model to terrestrial redox metabolism and could be a useful component to predicting biosignatures on other worlds.

6. Summary

We argue that life necessarily needs to perform redox chemistry to capture biomass.

- We point out that biochemicals exist overwhelmingly in a narrow range of redox states, as measured by a structural definition of redox (*Rr*) (Fig. 4).
- We argue that the narrow *Rr* distribution of actual, terrestrial biochemicals is the consequence of their need to be selected from the largest available phase space of possible chemicals. A large pool of chemicals is required because most chemicals will not satisfy the many constraints of biochemicals. We have shown that the space of possible chemicals is overwhelmingly concentrated in a relatively narrow range of redox states (Figs. 5 and 6).*
- Carbon—the central element in the structure of biochemicals—is likely to exist in a completely reduced or completely oxidized form in any planetary environment. This is in contrast to the carbon in biochemicals that is of intermediate *Rr* (Fig. 7). Therefore, life has to reduce oxidized carbon, or oxidize reduced carbon, in order to create molecules of intermediate redox state and, hence, assemble biochemical molecules and build biomass.

Based on the above three points, we argue that life has to carry out redox chemistry to capture carbon for biochemicals. Redox chemistry for carbon capture is distinct from life's use of redox chemistry to capture energy from environmental redox gradients. Redox gradients and their exploitation have been thoroughly discussed elsewhere. The use of redox chemistry to capture carbon requires the input of energy and may generate distinct waste products which, if volatile, could be useful atmospheric biosignatures.

Acknowledgments

We are very grateful to Dirk Schulze-Makuch (Washington State University), Tori Hoehler (NASA Ames), Stephen Freeland (NASA Astrobiology Institute, Hawaii) and Lewis Dartnell (University College London) for helpful comments and criticisms of past versions of the paper. We also acknowledge countless referees, whose previous harshly rejecting reports have strengthened the clarity of the final version of this paper.

*Note Added in Proof

We have recently been made aware of Kroll *et al.* (2011) who have analyzed the component chemicals in atmospheric haze in terms of their overall redox state. Kroll *et al.* also find that molecular diversity in haze components (C, H, and O-containing compounds) is overwhelmingly in compounds of intermediate redox state, and the end-product of environmental processes (*i.e.*, the most environmentally stable compounds) are ones of extreme oxidation state.

References

Alberts, B., Bray, D., Lewis, J., Martin, R., Roberts, K., and Watson, J.D. (1983) *Molecular Biology of the Cell*, Garland Publications, New York.

Allred, A.L. and Rochow, E.G. (1958) A scale of electronegativity based on electrostatic force. *Journal of Inorganic and Nuclear Chemistry* 5:264–268.

Bains, W. (2004) Many chemistries could be used to build living systems. *Astrobiology* 4:137–167.

Bairoch, A. (2000) The ENZYME database in 2000. *Nucleic Acids Res* 28:304–305.

Bard, A.J., Parsons, R., and Jordan, J. (1985) *Standard Potentials in Aqueous Solutions,* Marcel Dekker, New York.

Benner, S.A., Richardo, A., and Carrigan, M.A. (2004) Is there a common chemical model for life? *Curr Opin Chem Biol* 8: 672–689.

Bohacek, R.R., McMartin, C., and Guida, W.C. (1996) The art and practice of structure-based drug design. *Med Res Rev* 16:3–50.

Committee on the Origins and Evolution of Life, National Research Council. (2002) *Signs of Life: A Report Based on the April 2000 Workshop on Life Detection Techniques,* The National Academies Press, Washington DC.

Des Marais, D.J., Nuth, J.A., Allamandola, L.J., Boss, A.P., Farmer, J.D., Hoehler, T.M., Jakosky, B.M., Meadows, V.S., Pohorille, A., Runnegar, B., and Spormann, A.M. (2008) The NASA Astrobiology Roadmap. *Astrobiology* 8:715–730.

Entzeroth, M. (2003) Emerging trends in high-throughput screening. *Curr Opin Pharmacol* 3:522–529.

ExPasy. (2010) *Metabolic Map.* Retrieved July, 1, 2010, from http://www.expasy.ch/cgi-bin/show_thumbnails.pl.

Hann, M.M. and Oprea, T.I. (2004) Pursuing the leadlikeness concept in pharmaceutical research. *Curr Opin Chem Biol* 8:255–263.

Hitchcock, D.R. and Lovelock, J.E. (1967) Life detection by atmospheric analysis. *Icarus* 7:149–159.

Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T.J., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L., Altman, T., Paulsen, I., Keseler, I.M., and Caspi, R. (2010) Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* 11:40–79.

Keseru, G.M. and Makara, G.M. (2006) Hit discovery and hit-to-lead approaches. *Drug Discov Today* 11:741–748.

Kroll, J.H., Donahue, N.M., Jimenez, J.L., Kessler, S.H., Canagaratna, M.R., Wilson, K.R., Altieri, K.E., Mazzoleni, L.R., Wozniak, A.S., Bluhm, H., Mysak, E.R., Smith, J.D., Kolb, C.E., and Worsnop, D.R. (2011) Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol. *Nat Chem* 3:133–139.

Lehninger, A.L. (1972) *Bioenergetics,* W.A. Benjamin and Son, Menlo Park, CA.

Lipinski, C. and Hopkins, A (2004) Navigating chemical space for biology and medicine. *Nature* 432:855–861.

Lovelock, J.E. and Kaplan, I.R. (1975) Thermodynamics and the recognition of alien biospheres. *Proc R Soc Lond B Biol Sci* 189:167–181.

Martin, W. and Russell, M.J. (2007) On the origin of biochemistry at an alkaline hydrothermal vent. *Philos Trans R Soc Lond B Biol Sci* 362:1887–1925.

Metzler, D.E. (1977) *Biochemistry: The Chemical Reactions of Living Cells,* Academic Press, New York.

Morowitz, H., Kostelnik, J.D., Yang, J., and Cody, G.D. (2000) The origin of intermediary metabolism. *Proc Natl Acad Sci USA* 97:7704–7708.

Nealson, K. (1997) Sediment bacteria: who's there, what are they doing, and what's new? *Annu Rev Earth Planet Sci* 25:403–434.

Nelson, D.L. and Cox, M.M. (2009) *Lehninger: Principles of Biochemistry,* W.H. Freeman, New York.

Nielsen, K.F. and Smedsgaard, J. (2003) Fungal metabolite screening: database of 474 mycotoxins and fungal metabolites for dereplication by standardised liquid chromatography–UV–mass spectrometry methodology. *J Chromatogr A* 1002:111–136.

Oprea, T.I. and Marshall, G.R. (1998) Receptor-based prediction of binding affinities. *Perspectives in Drug Discovery and Design* 11:35–61.

Oprea, T.I., Davis, A.M., Teague, S.J., and Leeson, P.D. (2001) Is there a difference between leads and drugs? A historical perspective. *J Chem Inf Comput Sci* 41:1308–1315.

Pace, N.R. (2001) The universal nature of biochemistry. *Proc Natl Acad Sci USA* 98:805–808.

Pichersky, E. and Gang, D. R. (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Science* 5:439–445.

Polley, M.J., Burden, F.R., and Winkler, D.A. (2006) Simulation and modelling of chemical and biological complex systems. *Australian Journal of Chemistry* 59:859–864.

Proudfoot, J.R. (2002) Drugs, leads and drug-likeness: an analysis of some recently launched drugs. *Bioorg Med Chem Lett* 12:1647–1650.

Raulin, F., Bruston, P., Pailous, P., and Sternberg, R. (1995) The low temperature organic chemistry of Titan's geofluid. *Adv Space Res* 15:321–333.

Seager, S., Schrenk, M., and Bains, W. (2012) An astrophysical view of Earth-based metabolic biosignature gases. *Astrobiology* 12:61–82.

Shinbo, Y., Nakamura, Y., Altaf-Ul-Almin, M., Asahi, H., Kurokawa, K., Arita, M., Saito, K., Ohta, D., Shibata, D., and Kanaya, S. (2006) KNApSAcK: a comprehensive species-specific metabolic relationship database. In *Plant Metabolomics,* Biotechnology in Agriculture and Forestry Vol. 57, edited by K. Saito, R.A. Dixon, and L. Willmitzer, Springer, New York, pp 165–184.

Shock, E. (2009) Minerals as energy sources for microorganisms. *Econ Geol* 104:1235–1248.

Smith, E. and Morowitz, H.J. (2004) Universality in intermediary metabolism. *Proc Natl Acad Sci USA* 101:13168–13173.

Turnbull, A.P. and Swarbrick, M.E. (2009) Harnessing fragment-based drug discovery at CRT. *Drug Discovery World* 2009(Fall):57–64.

Warr, W.A. (1997) Combinatorial chemistry and molecular diversity: an overview. *J Chem Inf Comput Sci* 37:134–140.

Weininger, D. (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and coding rules. *J Chem Inf Comput Sci* 28:31–36.

Wunberg, T., Hendrix, M., Hillisch, A., Lobell, M., Meier, H., Schmeck, C., Wild, H., and Hinzen, B. (2006) Improving the hit-to-lead process: data-driven assessment of drug-like and lead-like screening hits. *Drug Discov Today* 11:175–180.

Address correspondence to:
Dr. William Bains
Rufus Scientific Ltd.
37 The Moor
Melbourn, Royston
Herts SG8 6ED
UK

E-mail: bains@mit.edu

Submitted 8 August 2011
Accepted 11 January 2012